TEMPORAL CONTROL OF ORGANOGENESIS

BY PHA-4/FOXA

by

Hui-Ting Hsu

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Oncological Sciences

The University of Utah

May 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Hui-Ting Hsu**

has been approved by the following supervisory committee members:

Susan Mango	, Chair	**12-23-2013** Date Approved
Don Ayer	, Member	**10-03-2013** Date Approved
Brad Cairns	, Member	**10-03-2013** Date Approved
Dean Tantin	, Member	**10-03-2013** Date Approved
Carl Thummel	, Member	**10-03-2013** Date Approved

and by **Brad Cairns and David Jones**, co-Chairs of

the Department/College/School of **Oncological Sciences**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

FoxA forkhead are conserved transcription factors that function in regulating foregut development of animals. FoxA factors regulate organogenesis from early cell fate specification to cell differentiation and functional morphogenesis by orchestrating timely transcriptional programs that drive these events at each developmental stage. Using the PHA-4/FoxA-dependent pharynx development in *C. elegans* as a model, we revealed that PHA-4 modulates temporal pharyngeal gene expression partly through DNA binding affinity. Pharyngeal genes with high affinity PHA-4 binding sites are competent to activate earlier when compared to genes with low affinity sites. We further demonstrated that affinity affects the level of PHA-4 occupancy at pharyngeal promoters and the decompaction of chromatin induced after PHA-4 binding. We tested the effect of affinity in response to changes in PHA-4 level and showed that temporal pharyngeal gene expression was sensitive to PHA-4 levels but was maintained properly within a range of PHA-4 level fluctuation. The dynamic level of PHA-4 during embryogenesis coordinates with the different binding affinity between PHA-4 and its targets to mediate the proper temporal order of pharyngeal gene expression.

Sequence-specific transcription factors are involved in regulating gene expression by recruiting RNA Pol II and/or mediating the release of paused RNA

Pol II at target genes. Poised Pol II at developmental regulated genes is shown to play an important role in controlling gene expression during embryogenesis in different organisms. However, poised Pol II is relatively rare in *C. elegans*. By examining the dynamic genome-wide Pol II occupancy at different embryonic stages, we discovered that poised Pol II was temporally regulated at different genes. Among poised pharyngeal genes, Pol II is found at genes that are either activated at later stages or highly expressed at earlier stages with a reduction in expression later. This result suggests that poised Pol II serves as a preparation for future gene expression or as a memory of past gene expression. Moreover, we showed that PHA-4 activity was required to load Pol II at poised pharyngeal genes. We propose that PHA-4/FoxA functions as a pioneer factor that primes gene activation by regulating Pol II.

In summary, PHA-4 utilizes DNA binding affinity by differentially binding its targets to control the timing of activation. The binding of PHA-4 also regulates Pol II occupancy at pharyngeal promoters. Our studies delineate the role of PHA-4 in regulating temporal gene expression during pharynx development and suggest that FoxA factors might function similarly in other systems.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Developmental processes, such as body axis patterning and organ formation, are driven by successive tiers of differential gene expression to regulate cell fate determination and cell differentiation. These events are often controlled by selector genes, which encode a special class of transcription factors functioning to control gene expression required for specifying cell, tissue, organ, and regional identity of animals (Mann and Carroll, 2002). The classic examples of selector genes are the Hox genes, which regulate the specification of body segments along anterior-posterior axes in many organisms. As Antonio Garcia-Bellido proposed, Hox genes regulate a cohort of downstream genes, also called realisators or effectors, which confer specific properties to cells involving cell division, cell adhesion, cell migration, and apoptosis to coordinate morphogenesis of body segments (Garcia-Bellido, 1975; Hueber and Lohmann, 2008).

Based on the features of selector genes, the loss of activity of selector genes often leads to the perturbation of the selector-regulated transcriptome and results in the loss of specific functional structures. An example manifested in *C. elegans* is pharynx development. The *C. elegans* pharynx is a pumping and feeding organ that consists of neurons, glands, structural cells, epithelia, valves and muscles. In forkhead transcription factor *pha-4/foxa* null mutants, *C. elegans* embryos completely fail to form a pharynx (Horner et al., 1998; Kalb et al., 1998; Mango et al., 1994). Conversely, ectopic PHA-4 expression throughout the embryo generates extra pharyngeal cells (Horner et al., 1998). Using a

temperature-sensitive mutant to investigate the effect of reducing PHA-4 activity at different developmental stages revealed that PHA-4 activity is essential for viability throughout life. Eliminating PHA-4 activity at early embryonic stages causes 100% lethality and no pharynx specification (Kiefer et al., 2007; Mango et al., 1994). Abrogating PHA-4 expression at later stages results in different degrees of pharyngeal defects, such as Pharynx unattached (Pun) and a stuffed pharynx (defects in pharyngeal pumping) phenotypes (Gaudet and Mango, 2002; Kiefer et al., 2007). These results indicate that PHA-4 is a master regulator and continuously required to regulate a comprehensive pharyngeal gene expression and form a functional pharynx during embryogenesis.

To ensure a reliable and reproducible developmental program, differential gene expression must be precisely regulated in a spatial and temporal manner, and proper gene expression levels must be maintained. A key question is how selector genes regulate the differential expression of multiple targets required for distinct developmental processes at specific stages. Selector genes often regulate differential gene expression through multiple strategies. For example, selector genes can function in a hierarchical manner as development proceeds by activating several genes at the top of a cascade. These targets directly activated by selectors will further carry out the effect of selectors through regulation of downstream genes (effectors) in the cascade. One example of hierarchical regulation is intestinal development in *C. elegans*. *C. elegans* GATA-type transcription factors *end-1* and *end-3* are essential for endoderm cell fate specification and activate a second set of GATA factors, ELT-2, ELT-4, and ELT-

7 (McGhee et al., 2007). ELT-2 and ELT-7 directly control the intestinal effector genes encoding the functional and structural proteins (McGhee et al., 2009; McGhee et al., 2007; Sommermann et al., 2010).

Other selector genes can directly bind multiple targets but regulate their expression differentially during development. For example, MyoD, a bHLH transcription factor belonging to the family of myogenic regulatory factors, globally regulates the transcriptome required for skeletal muscle development (Cao et al., 2010; Tapscott et al., 1988). MyoD directly regulates its targets by forming heterodimers with E-proteins (E12, E47, and HEB) and binding to the cis-element E-boxes (Puri and Sartorelli, 2000). MyoD controls the timing of target gene expression through feed-forward circuits and cooperative binding with co-factors, such as AP-1, Meis, Runx, and Sp1, to activate muscle-specific genes (Biesiada et al., 1999; Knoepfler et al., 1999; Sartorelli et al., 1990). MyoD also activates the expression of the transcriptional repressor RP58, which represses Inhibitor of DNA binding (Id) proteins in early differentiated muscle cells. Id2/Id3 are negative regulators of myogenesis that compete with MyoD to heterodimerize with E-proteins and therefore attenuate MyoD function. RP58 represses the expression of Id2/Id3 to make E-proteins available and thereby allows MyoD to activate the late differentiation genes (Benezra et al., 1990; Yokoyama et al., 2009). In summary, MyoD binds to various myogenic promoters and is sufficient to activate early myogenic genes but not late genes. On late genes, MyoD binding induces histone acetylation to facilitate binding of co-factors and other transcription factors, such as Myog and Pbx, to activate gene

expression (Cao et al., 2006). The time delay in the sequential recruitment of factors at MyoD late-stage targets assures the temporal expression patterns of late myogenic genes.

Functioning as a selector gene, PHA-4/FoxA broadly binds to the cis-element TRTTKRY (R= A/G, K= T/G, Y= T/C) within pharyngeal targets (and elsewhere in the genome) and activates their expression at different stages and in different pharyngeal cell types (Gaudet and Mango, 2002; Mango et al., 1994). Similar to MyoD, PHA-4 also regulates temporal pharyngeal gene expression through a feed-forward circuit. An example is the expression of pharyngeal muscle myosin, *myo-2*, which is activated by PHA-4 and a homeobox transcription factor, CEH-22 (Kalb et al., 1998; Okkema and Fire, 1994). PHA-4 initiates the expression of *ceh-22* in pharyngeal muscles during mid-embryogenesis (Mango et al., 1994; Vilimas et al., 2004). Both PHA-4 and CEH-22 bind to the *myo-2* promoter and activate *myo-2* expression at later embryonic stages (Ao et al., 2004; Gaudet and Mango, 2002; Zhong et al., 2010). Combinatorial regulation via feed-forward circuits is a common mechanism employed by selector genes to control differential gene expression.

Distinct from MyoD, PHA-4 also utilizes binding site affinity to control the onset of gene expression. It has been shown that PHA-4 binds to different TRTTKRY sequences found at endogenous targets with differential affinity (Gaudet and Mango, 2002). Furthermore, PHA-4 binding affinity modulates the temporal expression of pharyngeal genes. Changing a PHA-4 binding site to increase or decrease the binding affinity within the promoter context resulted in

advanced or delayed pharyngeal gene expression (Gaudet and Mango, 2002). This result indicates that PHA-4 binding site affinity is a critical determinant of temporal pharyngeal gene expression. In this chapter, I will discuss that PHA-4/FoxA, as a model selector gene, orchestrates temporal gene expression to govern pharynx development. A proposed model of PHA-4/FoxA-regulated temporal pharyngeal gene expression will be described in detail. The connection and a potential role of PHA-4/FoxA in regulating RNA Polymerase II will be discussed as well.

1.2 Temporal control of pharyngeal organogenesis by PHA-4/FoxA

1.2.1 Pharynx development in *C. elegans* and PHA-4 expression

To form a pharynx, *C. elegans* early embryos first generate a group of pharyngeal precursor cells derived from two cell lineages, EMS and AB (2 blastomeres in the 4-cell embryo). The number of pharyngeal precursors is increased during gastrulation. At mid-embryogenesis, the pharyngeal precursors assemble and form the pharyngeal primordium. The pharyngeal precursors subsequently specify six different pharyngeal cell types, including pharyngeal muscle, gland, valve, epithelial, neural, and structural cells. Morphogenesis and terminal differentiation of the different pharyngeal cell types further establish a functional pharynx (Figure 1.1) (Mango, 2009).

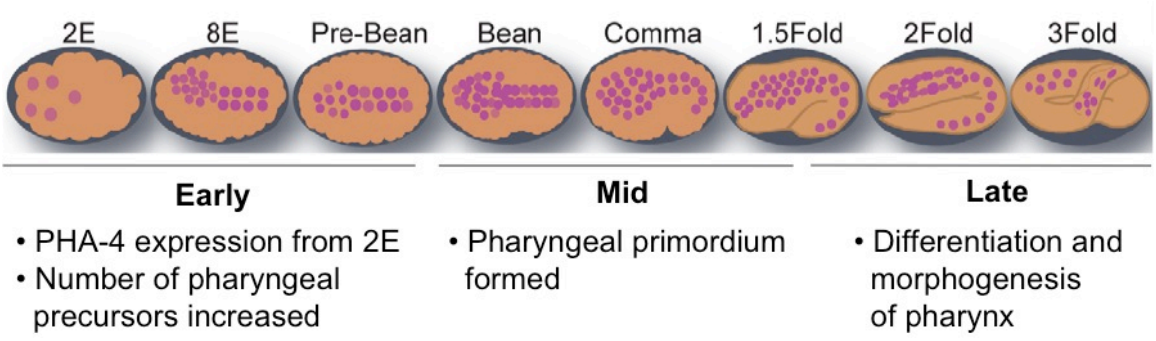


Figure 1.1
A cartoon of PHA-4 expression and pharynx development during embryogenesis. PHA-4 is activated from the 2E stage in the pharyngeal precursors. At early embryonic stages, the pharyngeal precursors divide during gastrulation. The pharyngeal primordium is formed and pharyngeal cell types are subsequently specified during mid-embryogenesis when embryonic cell division is completed. A functional pharynx is formed after cell differentiation and morphogenesis. (Adopted from Fakhouri et al., 2010)

Prior to gastrulation, maternally contributed Notch signaling initiates two separate cascades to activate *pha-4* expression in the pharyngeal precursors derived from EMS and AB descendants (Mango et al., 1994; Priess et al., 1987). PHA-4 is the earliest marker of pharyngeal expression. PHA-4 is activated in all pharyngeal cells during early gastrulation, at the 2E stage (the stage when embryos contain two endodermal cells, in total ~28 cells in the embryo), and maintained throughout life (Mango, 2009). PHA-4 expression is increased during embryogenesis and reaches its maximum level during mid-embryogenesis (Horner et al., 1998; Smith and Mango, 2007). PHA-4 is also expressed in midgut and hindgut, but its function in these tissues is not well characterized (Azzaria et al., 1996; Horner et al., 1998; Updike and Mango, 2007). In larvae, PHA-4 is expressed in the somatic gonad, where it is important for its formation (Azzaria et al., 1996; Updike and Mango, 2007).

1.2.2 PHA-4 is an organ selector gene for pharynx development

PHA-4 is the central regulator of pharynx development and *pha-4* null alleles delete virtually the entire pharynx (Mango et al., 1994). This dramatic phenotype reveals that PHA-4 is an organ selector gene for foregut formation in worms, and its activity is globally required for all pharyngeal expression, including early developmental regulators that help specify pharyngeal cell types, as well as late differentiation genes encoding structural proteins and enzymes. Therefore, losing PHA-4 activity at different embryonic stages results in a wide spectrum of pharyngeal phenotypes. Eliminating PHA-4 expression at early embryonic stages

causes embryonic lethality, and nascent pharyngeal cells are transformed to other cell fates. Withdrawing PHA-4 activity at later stages leads to different degrees of pharyngeal functional defects such as irregular pharyngeal pumping and misplaced cells (Gaudet and Mango, 2002; Kaltenbach et al., 2005; Kiefer et al., 2007; Mango et al., 1994).

How PHA-4/FoxA modulates transcription to control pharynx development is one of the key questions for understanding the function of selector genes. One possibility is that PHA-4 regulates transcription by altering the chromatin environment. Studies of PHA-4 mammalian orthologs, the FoxA factors, demonstrate that FoxA transcription factors function as pioneer factors (Cirillo et al., 2002; Lupien et al., 2008). Pioneer factors initiate binding to targets prior to transcription, before other factors can access DNA. The binding of pioneer factors induces an open chromatin conformation that allows other factors to bind (Zaret and Carroll, 2011). In the initial model, FoxA blocked H1 histones, leading to local decompaction (Zaret and Carroll, 2011), but more recent studies have suggested FoxA binding leads to loss or destabilization of a nucleosome (Li et al., 2012). On the other hand, FoxA factors in mammals and Fkh factors in yeast can have long-range effects that promote transcription factors loading or recombination (Lupien et al., 2008; Sun et al., 2002). In *C. elegans* pharyngeal nuclei, PHA-4 bound to extrachromosomal arrays carrying PHA-4 targets prior to pharyngeal transcription and resulted in broad chromatin decompaction of the arrays (Fakhouri et al., 2010). The chromatin decompaction induced by PHA-4 activity is partly accomplished through recruitment of the histone variant HTZ-

1/H2A.Z to a subset of pharyngeal promoters (Updike and Mango, 2006). Genome-wide studies in various organisms indicate that H2A.Z containing nucleosomes flanked around transcription start sites (Barski et al., 2007; Luk et al., 2010). The presence of HTZ-1/H2A.Z also correlates with the occupancy of RNA Pol II, but not with actual transcription (Whittle et al., 2008). These observations suggest that PHA-4 might prime pharyngeal expression by affecting the chromatin environment and facilitating recruitment of downstream factors.

1.2.3 PHA-4 regulates temporal pharyngeal gene expression

Recent genome-wide chromatin immunoprecipitation-sequening (ChIP-Seq) studies have shown that PHA-4 binds more than 5000 genomic regions in mixed-stage embryos (Zhong et al., 2010). 90% of PHA-4 binding sites were assigned to protein coding genes that gave rise to approximately 4000 PHA-4 direct targets, including the previously identified PHA-4 targets (Ao et al., 2004; Gaudet and Mango, 2002; Gaudet et al., 2004). RNA-Seq further confirmed that 87% of PHA-4 targets were actively expressed in embryos. These results indicate that PHA-4 directly binds its targets to broadly activate gene expression during embryogenesis.

Previous studies revealed that PHA-4 alone functions as a weaker transcriptional activator by which three PHA-4 binding sites activated weak and patchy pharyngeal expression of a green fluoresce protein (GFP) reporter (Gaudet et al., 2004). The combination of PHA-4 sites with additional cis-regulatory elements supported more stable and specific pharyngeal gene

expression either in the whole pharynx or in specific pharyngeal cell types (Gaudet et al., 2004; Okkema and Fire, 1994; Raharjo and Gaudet, 2007). Conversely, pharyngeal cis-regulatory regions without PHA-4 sites failed to activate pharyngeal expression of GFP reporters (Gaudet and Mango, 2002; Raharjo et al., 2010). This result indicates that PHA-4 is required but is not sufficient to support robust pharyngeal gene expression.

A fundamental question is how PHA-4 differentially regulates its targets that function in various processes of pharynx development. The features of PHA-4 acting as a pioneer factor and the necessity of PHA-4 binding sites in modulating pharyngeal activation indicate that the interaction of PHA-4 with its target DNA is the primary determinant of pharyngeal expression. Indeed, the binding affinity between PHA-4 and its target promoters is a key input to regulate proper temporal pharyngeal gene expression. Previous studies from the Mango lab showed that pharyngeal promoters containing high affinity PHA-4 binding sites tend to activate earlier in development compared to promoters with low affinity PHA-4 sites (Gaudet and Mango, 2002). More importantly, the presence of a point mutation to change PHA-4 binding affinity within the promoter shifted the timing of pharyngeal activation—higher affinity activated expression earlier whereas lower affinity led to later onset (Gaudet and Mango, 2002). The mutations did not alter the strength of expression but only the onset of timing. These results reveal how the regulatory information embedded in the genome plays a role in setting up differential gene expression.

PHA-4 binding site affinity cannot fully account for all temporal pharyngeal gene expression. One example is the expression of pharyngeal muscle myosin, *myo-2*, which contains high affinity PHA-4 binding sites but is not activated until late embryogenesis. As mentioned previously, detailed promoter assays revealed that both PHA-4 and CEH-22 are required to activate *myo-2* (Gaudet and Mango, 2002; Mango et al., 1994; Okkema et al., 1997). These studies indicate a feed-forward mechanism of *myo-2* expression in which PHA-4 first activates *ceh-22* at mid embryogenesis, and then works combinatorily with CEH-22 at the *myo-2* promoter, to activate *myo-2* at later stages. The combination of binding site affinity and other factors in controlling precise pharyngeal gene transcription motivated us to search for other potential factors and cis-elements that might work with PHA-4 to refine the patterns of pharyngeal gene expression.

To identify the regulatory elements that function together with PHA-4 to mediate temporal pharyngeal gene expression, pharyngeal specific targets were grouped based on the timing of their onset (early versus late pharyngeal genes). Bioinformatic analyses were conducted to search for overrepresented motifs specifically in early or late pharyngeal targets, termed early or late elements (Gaudet et al., 2004). Moreover, synthetic promoters confirmed that these identified temporal elements functioned as pharyngeal enhancers and contributed to temporal control of pharyngeal expression (Gaudet et al., 2004). These temporal motifs, including GAGA rich sequences, were highly enriched in genome-wide PHA-4 binding targets identified from ChIP-Seq (Zhong et al.,

2010). These data provide a framework for understanding the logic of temporal control of pharyngeal gene expression.

Based on previous studies, we proposed a model for PHA-4-controlled temporal pharyngeal gene expression in which PHA-4 binds to sites with different affinity and functions in combination with factors associated with “early elements” or “late elements” to fine-tune the onset of pharyngeal gene expression (Figure 1.2). In the proposed model, PHA-4 affinity binding sites are differentially bound by PHA-4 as PHA-4 expression levels change during embryogenesis. PHA-4 binds to high affinity sites efficiently to initiate gene expression when the PHA-4 level is low at early embryonic stages. Low affinity targets are not bound stably to activate transcription until the PHA-4 level is increased later in development. In addition, other transcription factors express at different embryonic stages and bind to early or late elements at pharyngeal promoters/enhancers. The availability of both inputs from PHA-4 and secondary factors determines the proper pharyngeal expression patterns. In summary, high affinity PHA-4 sites and early elements promote early onset of pharyngeal expression in contrast to low affinity PHA-4 sites and late elements, which delay gene activation. The matching of PHA-4 affinity sites and different temporal elements at individual pharyngeal promoters/enhancers controls the precise timing of pharyngeal activation.

The proposed PHA-4 affinity model would predict that pharyngeal gene expression is sensitive to PHA-4 levels. Increased PHA-4 expression might saturate the PHA-4 binding sites and advance the onset of pharyngeal gene

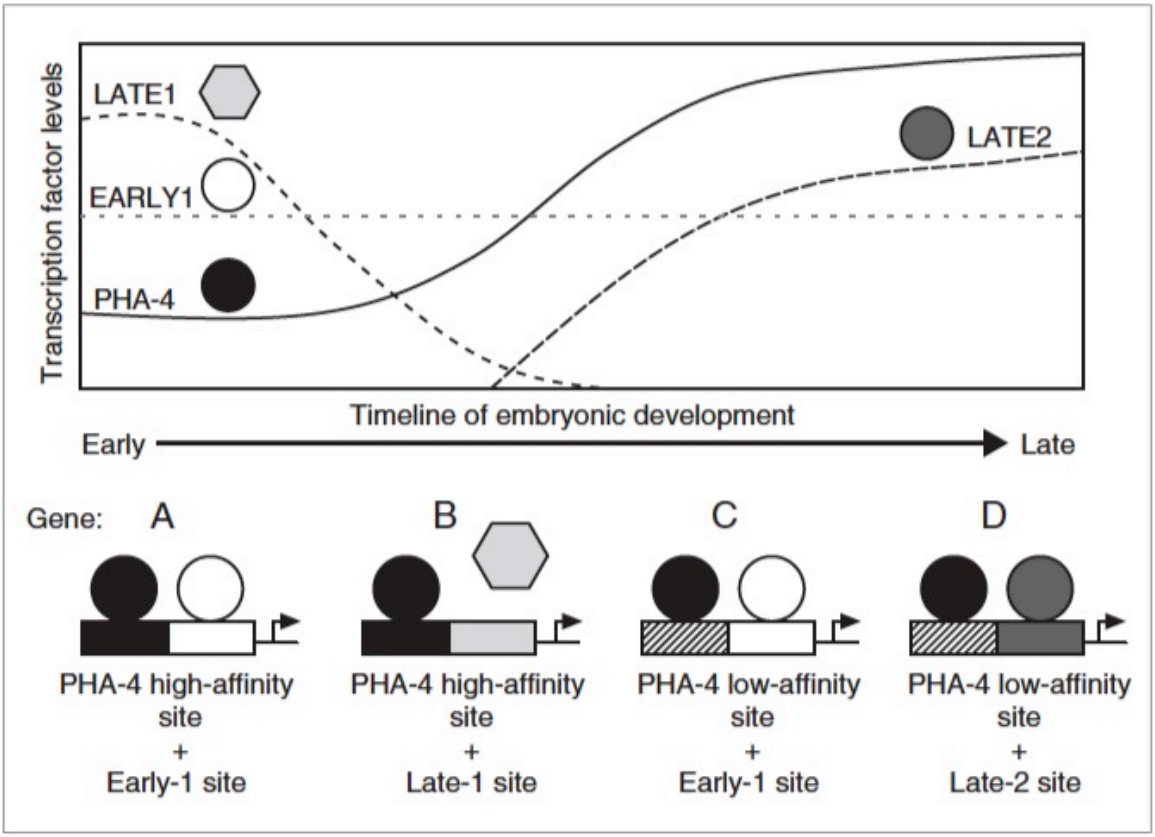


Figure 1.2
A model for temporal control of pharyngeal gene expression in *C. elegans*. Four transcription factors including PHA-4 are differentially expressed during embryogenesis. Gene A, B, C, and D are four pharyngeal genes express from different stages. In this model, both PHA-4 and other factors are required for proper pharyngeal expression. During early embryonic stages, PHA-4 level is low and only high affinity PHA-4 sites (Gene A and B) bound by PHA-4 are efficient for activating expression in comparison to low affinity PHA-4 sites (Gene C and D). Other transcription factors, which could be an activator or a repressor (white and gray circle and light gray hexagon), associate with early and late elements at different stages depend on their availability and levels. The onset of pharyngeal expression is determined by the combinatorial inputs from both PHA-4 and other transcription factors. Gene A, which contains a high affinity PHA-4 site, and an early element is activated by earliest time point by PHA-4 and the early factor. The delayed recruitment of late factors associated with late elements retards the timing of expression. The precise timing of pharyngeal expression is fine tuned by the match of different PHA-4 affinity sites and temporal elements (Adopted from Banerjee and Slack, 2005).

expression. This assumption predicts that temporal patterns of pharyngeal expression could be maintained within a range of PHA-4 levels in which the difference of PHA-4 binding to various affinity sites still translates to the differential timing in activation of gene expression. In other words, dramatic fluctuation of PHA-4 levels beyond the tolerance of binding affinity will result in misregulation of pharyngeal gene expression. We later tested this idea of affinity-controlled differential gene expression, which is discussed in Chapter 2.

1.3 RNA Polymerase II pausing and differential gene regulation

How do selector genes regulate gene expression after binding to their targets? The major target in controlling transcription of protein-coding genes is RNA Pol II. The transcriptional cycle of RNA Polymerase II contains at least 3 different phases: (1) initiation, where Pol II is recruited along with general transcription factors (GTFs) to form the preinitiation complex (PIC) at a promoter and starts to synthesize short RNA transcripts, (2) elongation, when Pol II escapes from the initiation site and moves into gene bodies to produce RNA, and (3) termination, in which Pol II and RNA transcripts disassociate from the DNA when transcription is completed (Adelman and Lis, 2012; Gilchrist and Adelman, 2012). It was long thought that sequence-specific transcription factors control gene expression mainly at the transcription initiation step by recruiting the transcription machinery to specific loci. Certain unusual cases were known, such as the *Drosophila* heat shock gene, in which heat shock factor (HSF) regulated

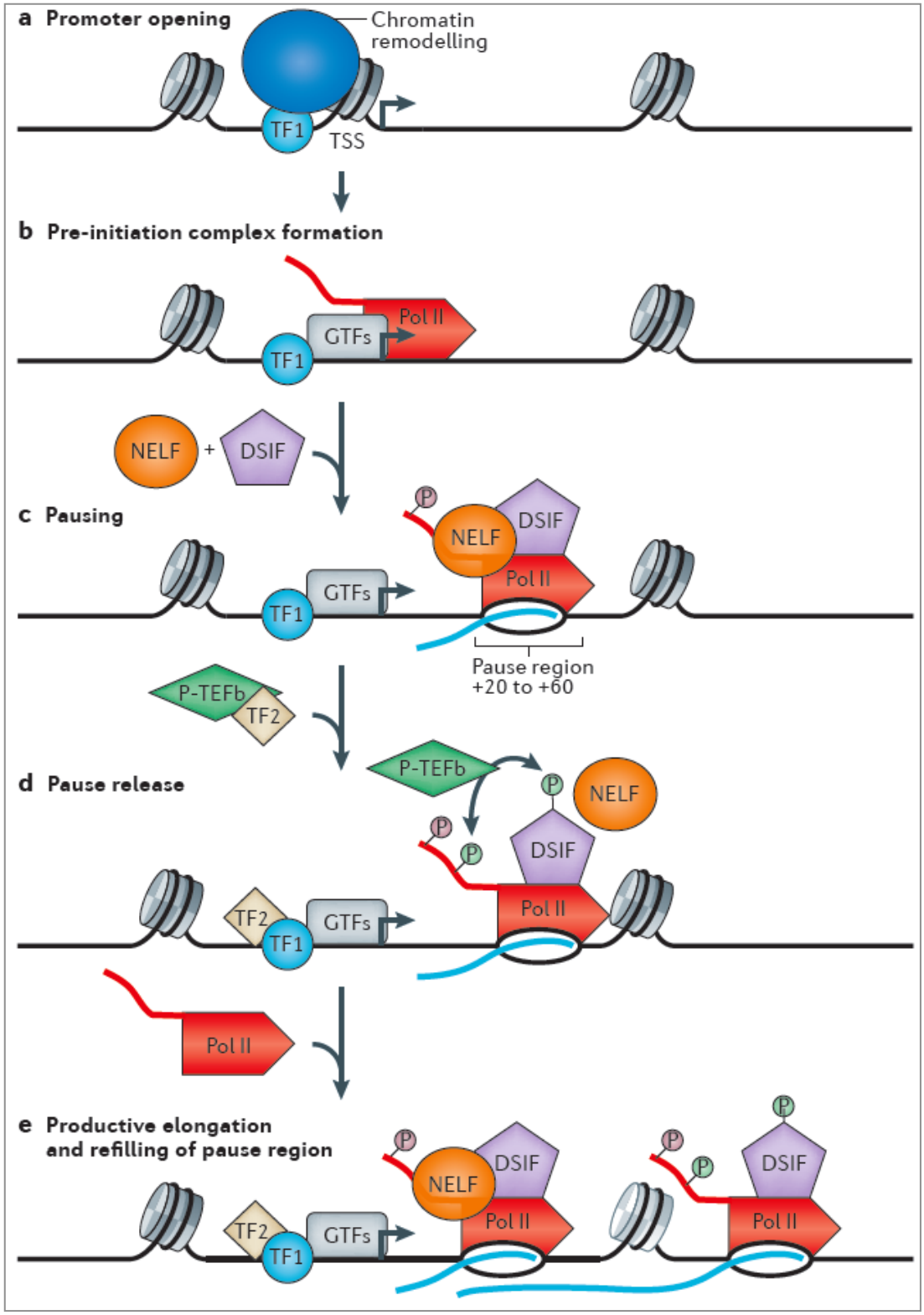
the release of promoter proximal paused Pol II at *hsp70* upon heat shock stimulus (O'Brien and Born, 1991). However, Pol II pausing was thought to be rare. Recent genome-wide studies of Pol II occupancy have revealed widespread Pol II pausing at promoter proximal regions in *Drosophila*, mouse, and human cells (Core et al., 2008; Muse et al., 2007; Nechaev et al., 2010; Rahl et al., 2010; Zeitlinger et al., 2007). These results suggest another layer of gene expression regulation postinitiation of RNA Pol II. In Chapter 3, I examined Pol II occupancy and PHA-4 in *C. elegans*. Below I summarize what is known about paused Pol II.

1.3.1 Discovery of Pol II pausing at promoter proximal regions

Depending on the cell type and organism, it has been shown that 10–60% of genes exhibit 5' enriched Pol II signals at promoter proximal regions (Core et al., 2012; Nechaev et al., 2010; Zeitlinger et al., 2007). Further studies showed that the 5' enriched Pol II is transcriptionally engaged, with phosphorylation of Ser5 with the Pol II carboxyl-terminal domain repeat (CTD), but pauses downstream of the transcriptional start site after generating 30 to 50-nucleotide long RNA transcripts (Figure 1.3) (Core et al., 2012). Other studies have shown Pol II at the 5' end of genes but have not determined its phosphorylation status or whether it is actively engaged in transcribing; I refer to this as poised Pol II (Adelman and Lis, 2012). In *Drosophila* early embryos, many developmental regulators, such as Hox genes, contain poised Pol II prior to gene activation

Figure 1.3

Establishment and release of promoter proximal paused Pol II. (a) Promoter opening is often associated with binding of sequence-specific transcription factors (shown as TF1) and recruiting chromatin remodelers (blue oval) to create an accessible chromatin environment for the assembly of transcriptional machinery. (b) The preinitiation complex (PIC) containing general transcription factors and RNA polymerase II is recruited to a promoter region. Sequence-specific transcription factors also involve in regulating the recruitment of PIC. (c) Pol II pausing after transcriptional initiation involves the association of 5,6-dichloro-1-b-D-ribofuranoxyl-benzimidazole sensivitiy-inducing factor (DSIF) and negative elongation factor (NELF). The paused Pol II is phophorylated on Ser5 of its CTD. (d) Pause release is regulated by the recruitment of p-TEFb directly or indirectly through sequence-specific transcription factors (shown as TF2). p-TEFb phosphorylates DSIF, NELF, and the Pol II CTD to release paused Pol II. (e) After paused Pol II is released to productive elongation, another Pol II complex might form at promoter region for efficient RNA expression (Adopted from Adelman and Lis, 2012).


a Promoter opening
Chromatin remodelling
TF1
TSS
b Pre-initiation complex formation
Pol II
GTFs
TF1
NELF
+
DSIF
c Pausing
P
NELF
DSIF
Pol II
GTFs
TF1
Pause region
+20 to +60
P-TEFb
TF2
d Pause release
P-TEFb
P
NELF
DSIF
Pol II
TF2
GTFs
TF1
Pol II
e Productive elongation and refilling of pause region
P
NELF
DSIF
Pol II
TF2
GTFs
TF1
P
DSIF
Pol II

(Zeitlinger et al., 2007). These data suggest that Pol II poising is an important regulatory step in transcriptional regulation.

1.3.2 Mechanism of promoter proximal Pol II pausing

Several DNA sequence motifs and sequence-specific binding factors are involved in setting up Pol II pausing. Sequence analysis of paused genes in *Drosophila* embryos indicated that paused promoters tend to be more GC-rich in sequence (Hendrix et al., 2008). Several sequence motifs are overrepresented in paused promoters, such as the downstream promoter element (DPE; RGWYV(T)), the initiator motif (Inr motif; YYANWYY), the pausing button (PB; KCGRWCG), and the GAGA motif (Hendrix et al., 2008; Lee et al., 2008). Among these motifs, the GAGA motif is bound by GAGA factor (GAF), which encodes a Trithorax-like protein and alters the chromatin environment via disrupting nucleosome assembly at the core promoters in favor of Pol II occupancy (Gilchrist et al., 2010). The nucleosome-free chromatin of the paused promoters results in long Pol II dwelling and allows DSIF and NELF to bind to the nascent transcripts and form a stable paused Pol II complex at promoter proximal regions (Missra and Gilmour, 2010). However, GAF-associated Pol II pausing only accounts for 20% of paused genes in *Drosophila*. This result suggests that other factors are involved in regulating Pol II pausing.

A recent study found that a previously identified core promoter element, Motif 1 (Ohler et al., 2002), was enriched in the paused genes lacking the GAGA motif (Li and Gilmour, 2013). Using affinity chromatography, Li further identified a

novel, ubiquitously expressed zinc finger protein specifically bound to Motif 1 in *Drosophila* embryos, named Motif 1 binding protein (M1BP). Different from GAF-mediated strong Pol II pausing, M1BP-bound genes showed a lower level of Pol II pausing (Li and Gilmour, 2013). Gene ontology analysis indicated that M1BP bound genes are enriched for basic cellular processes and exhibited more uniform expression during development. In contrast, GAGA factor is associated with Pol II pausing at developmentally regulated genes (Li and Gilmour, 2013). The difference in gene function and expression patterns as well as the level of paused Pol II suggests that the mechanism of M1BP-mediated Pol II pausing is distinct from that of GAF. In *Drosophila*, these two factors are responsible for ~50% of paused genes, which implies that other factors might be involved in setting up pausing at genes in different regulatory programs.

1.3.3 Regulation of gene expression through Pol II pausing

The accumulation of Pol II at promoter proximal regions suggests that the expression of these genes is regulated by releasing paused Pol II to promote productive transcriptional elongation, which is associated with phosphorylation of Serine 2 (Ser2) within the C terminal domain (CTD) of Pol II (Hirose and Ohkuma, 2007; Ni et al., 2004). The release of paused Pol II relies on the activity of P-TEFb (positive transcription elongation factor b) to phosphorylate its substrates including Pol II, DSIF, and NELF (Wada et al., 1998; Yamaguchi et al., 2013; Yamaguchi et al., 1999). This observation indicates that regulation of P-TEFb is a central node to control expression of poised genes. Several

sequence-specific transcription factors, for example c-Myc, NF-κB, MyoD, and estrogen receptor, have been shown to recruit P-TEFb to activate its targets through direct interaction with P-TEFb (Barboric et al., 2001; Rahl et al., 2010; Simone et al., 2002; Welboren et al., 2009). P-TEFb is also a component of complexes including the super elongation complex (SEC) (Luo et al., 2012). SEC is a massive complex containing a large number of proteins called Mediators, which provide a scaffold to bridge protein-protein interactions in regulating transcription (Luo et al., 2012; Takahashi et al., 2011). Therefore, sequence-specific transcription factors could directly or indirectly regulate transcription elongation through interacting with SEC.

1.3.4 Function of Pol II pausing in development

Pol II pausing has been implicated as an important checkpoint in regulating gene expression during development. In *Drosophila* embryos, Pol II is poised at a subset of genes that are inactive in certain cell types and tissues, such as *brinker* (*brk*), which is expressed in neurectoderm but inactive and poised in mesodermal cells (Zeitlinger et al., 2007). In contrast, Pol II is also poised at another set of developmentally regulated genes that are actively expressed at later developmental time points. These observations suggest two developmental functions for Pol II poising. Poised Pol II could be an indication of transcriptional repression and also reflect a memory of prior transcriptional activity. On the other hand, poised Pol II could act as a preparation of gene activation at later developmental stages.

The poising of Pol II at developmental regulator genes expressed at later stages suggests a preparation for rapid gene activation to execute developmental decisions. The fast induction of poised genes has been reported from studies of *Drosophila hsp70* (Wilkins and Lis, 1997). In addition, poised genes tend to activate in a synchronous manner compared to nonpoised genes, which are more variable and stochastic, as determined by in situ hybridization in *Drosophila* embryos (Boettiger and Levine, 2009). Switching poised genes to nonpoised or weakly poised promoters resulted in disrupting the synchronous patterns of gene expression and caused morphogenetic defects (Lagha et al., 2013). This study suggests that the rapid and synchronous activation of poised genes is crucial to coordinate temporal gene expression required for specific developmental processes.

On the other hand, Pol II poising can function to maintain an open chromatin configuration. The long dwelling time of poised Pol II at promoter proximal regions inhibits nucleosome assembly at the core promoters; therefore maintaining these promoters in an open conformation. This idea is supported by several studies. The loss of paused Pol II by reducing NELF leads to increased nucleosome occupancy and silencing of gene expression at several highly poised genes (Gilchrist et al., 2008). Comparing the nucleosome profiles between poised and nonpoised promoters showed a significant correlation of Pol II poising and nucleosome deprivation around the transcription start site (TSS) (Gilchrist et al., 2010). Keeping promoters in an open chromatin configuration may facilitate rapid or synchronous induction of gene activation and the recycling

of Pol II complex to re-enter the promoters for the next round of transcription, for example in response to cell signaling. Overall, Pol II poising provides a layer of control in the regulation of gene expression required for robustness of development.

1.3.5 Pol II poising in *C. elegans*

To date, the most pronounced Pol II poising in *C. elegans* has been identified in young larvae (L1) arrested in development by lack of food (Baugh et al., 2009). In starved L1 larvae, Pol II poised specifically at growth and developmental genes, whose expression was repressed in response to starvation but later resumed expression when food became available (Baugh et al*.,* 2009). This study indicates that Pol II poising reflects the memory of expression, but perhaps also the priming for future transcription in response to feeding.

A recent study might explain how Pol II pausing is regulated during L1 starvation. During starvation or under environmental stress, an unknown signal can induce the recruitment of the nuclear protein ZFP-1 and histone methyltransferase DOT-1.1 to widely-expressed essential genes (Cecere et al., 2013). These genes included the growth genes previously reported to be paused during L1 starvation (Baugh et al*.,* 2009). The recruitment of the ZFP-1/DOT1.1 complex resulted in the increase of methylation of histone H3 at position 79 (H3K79), which was associated with slowing Pol II and poising at promoters (Cecere et al., 2013). This study revealed a negative feedback mechanism to

mediate gene expression through Pol II pausing in response to environmental stimuli. This result also highlights the connection of histone modification in regulating Pol II dynamics, which might provide a framework to unravel the molecular mechanism of Pol II pausing.

Distinct from *Drosophila*, Pol II pausing in *C. elegans* embryos is relatively rare (Kruesi et al., 2013; Zhong et al., 2010). Pol II ChIP-Seq performed by modENCODE identified ~2% of genes exhibited 5' stalled Pol II in mixed-stage embryos (Zhong et al., 2010). A recent publication from the Meyer and Lis labs further indicated that few genes, only 0.38% of genes in embryos, were associated with transcriptionally engaged and paused Pol II using Global Run On Sequencing (GRO-Seq) (Kruesi et al., 2013). This phenomenon might be explained by the lack of several pausing factors from *C. elegans*, such as GAGA factor and NELF. The mechanism of Pol II poising in arrested L1 larvae mentioned above might be distinct from the mechanisms to regulate Pol II poising in embryos.

Studies that have combined RNA-Seq and Pol II ChIP-Seq have suggested that Pol II occupancy is dynamic. In *C. elegans* arrested L1 larvae, Pol II was poised at growth genes during starvation. Within an hour after feeding, Pol II occupancy at promoters decreased as genes responding to food and started to express (Baugh et al., 2009). This is similar to studies with *Drosophila* in which Pol II was dynamically regulated during embryogenesis (Gaertner et al., 2012). Genes lacked poised Pol II at early stages and gained Pol II prior to transcription. Intriguingly, poised Pol II was found in multiple tissues, including those that would

never transcribe the particular genes. Thus, in *Drosophila* embryos, poised Pol II reflects temporal but not cell type-specific regulation (Gaertner et al., 2012). These data suggest that Pol II poising is dynamically regulated and stage-specific. Therefore, careful examination of Pol II poising in *C. elegans* embryos requires using synchronized populations.

1.4 Pharynx development provides a model to study temporal gene expression

Pharynx development in *C. elegans* is driven by temporal activation of subsets of pharyngeal genes at each developmental stage. This process is tightly regulated by PHA-4 activity. Therefore, pharynx development is a good model to explore different mechanisms in controlling temporal gene expression during development, including Pol II poising. In fact, the feature of PHA-4 function as a pioneer factor makes it appealing to test whether PHA-4 modulates temporal pharyngeal activation through Pol II poising. The PHA-4 activity to induce chromatin decompaction might facilitate the recruitment of Pol II or reflect the exchange of a nucleosome for Pol II. These ideas have not been tested nor have pioneer factors been reported to be involved in Pol II poising.

The motivation of this study is to better understand how selector genes contribute to temporal gene expression in controlling complex developmental processes, such as organ formation. In this thesis, I focused on studying events that occur upstream of temporal gene expression by examining PHA-4 binding to various pharyngeal promoters and their transcriptional outcomes (Chapter 2).

Next, I examined patterns of genome-wide Pol II occupancy at pharyngeal genes (Chapter 3). Results from my studies indicated that DNA binding site affinity affects the level of PHA-4 occupancy at target promoters to modulate the onset of pharyngeal gene expression. PHA-4-controlled temporal pharyngeal gene expression is partly through regulating RNA Pol II occupancy at pharyngeal promoters.

1.5 References

Adelman, K., and Lis, J. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Ao, W., Gaudet, J., Kent, W.J., Muttumu, S., and Mango, S.E. (2004). Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. Science *305*, 1743–1746.

Azzaria, M., Goszczynski, B., Chung, M.A., Kalb, J.M., and McGhee, J.D. (1996). A fork head/HNF-3 homolog expressed in the pharynx and intestine of the Caenorhabditis elegans embryo. Dev. Biol. *178*, 289–303.

Banerjee, D. and Slack, F.J. (2005). Temporal and spatial patterning of an organ by a single transcription factor. Genome Biol. *6,* 205.

Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Mol. Cell *8*, 327–337.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Baugh, L., Demodena, J., and Sternberg, P. (2009). RNA Pol II accumulates at promoters of growth genes during developmental arrest. Science *324*, 92–94.

Benezra, R., Davis, R.L., Lassar, A., Tapscott, S., Thayer, M., Lockshon, D., and Weintraub, H. (1990). Id: a negative regulator of helix-loop-helix DNA binding proteins. Control of terminal myogenic differentiation. Ann. N. Y. Acad. Sci. *599*, 1–11.

Biesiada, E., Hamamori, Y., Kedes, L., and Sartorelli, V. (1999). Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac alpha-actin promoter. Mol. Cell. Biol. *19*, 2577–2584.

Boettiger, A., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471–473.

Cao, Y., Kumar, R.M., Penn, B.H., Berkes, C.A., Kooperberg, C., Boyer, L.A., Young, R.A., and Tapscott, S.J. (2006). Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters. EMBO J. *25*, 502–511.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G., Parker, M., MacQuarrie, K., Davison, J., Morgan, M., Ruzzo, W., et al. (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev. Cell *18*, 662–674.

Cecere, G., Hoersch, S., Jensen, M., Dixit, S., and Grishok, A. (2013). The ZFP-1(AF10)/DOT-1 complex opposes H2B ubiquitination to reduce Pol II transcription. Mol. Cell *50*, 894–907.

Cirillo, L., Lin, F., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Core, L., Waterfall, J., Gilchrist, D., Fargo, D., Kwak, H., Adelman, K., and Lis, J. (2012). Defining the status of RNA polymerase at promoters. Cell Rep. *2*, 1025–1035.

Core, L., Waterfall, J., and Lis, J. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Fakhouri, T.H., Stevenson, J., Chisholm, A.D., and Mango, S.E. (2010). Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1001060

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep. *2*, 1670–1683.

Gaudet, J., and Mango, S. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821–825.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in Drosophila. Ciba Foundation Symp. 0,161–182.

Gaudet, J., Muttumu, S., Horner, M., and Mango, S. (2004). Whole-genome analysis of temporal gene expression during foregut development. PLoS Biol. *2*.

Gilchrist, D., and Adelman, K. (2012). Coupling polymerase pausing and chromatin landscapes for precise regulation of transcription. Biochim. Biophys. Acta *1819*, 700–706.

Gilchrist, D., Dos Santos, G., Fargo, D., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540–551.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. *22*, 1921–1933.

Hendrix, D., Hong, J.-W., Zeitlinger, J., Rokhsar, D., and Levine, M. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc. Natl. Acad. Sci. USA. *105*, 7762–7767.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. J. Biochem. *141*, 601–608.

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., and Mango, S.E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. Genes Dev. *12*, 1947–1952.

Hueber, S., and Lohmann, I. (2008). Shaping segments: Hox gene function in the genomic age. BioEssays *30*, 965–979.

Kalb, J., Lau, K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P., and McGhee, J. (1998). pha-4 is Ce-fkh-1, a fork head/HNF-3alpha,beta,gamma homolog that functions in organogenesis of the C. elegans pharynx. Development *125*, 2171–2180.

Kaltenbach, L.S., Updike, D.L., and Mango, S.E. (2005). Contribution of the amino and carboxyl termini for PHA-4/FoxA function in Caenorhabditis elegans. Dev. Dyn. *234*, 346–354.

Kiefer, J., Smith, P., and Mango, S. (2007). PHA-4/FoxA cooperates with TAM-1/TRIM to regulate cell fate restriction in the C. elegans foregut. Dev. Biol. *303*, 611–624.

Knoepfler, P.S., Bergstrom, D.A., Uetsuki, T., Dac-Korytko, I., Sun, Y.H., Wright, W.E., Tapscott, S.J., and Kamps, M.P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic Acids Res. *27*, 3752–3761.

Kruesi, W., Core, L., Waters, C., Lis, J., and Meyer, B. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. eLife *2*.

Lagha, M., Bothma, J., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D., Zeitlinger, J., et al. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell *153*, 976–987.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol. Cell. Biol. *28*, 3290–3300.

Li, J., and Gilmour, D.S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J. *32*, 1829–1841.

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K. (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell *151*, 1608–1616.

Luk, E., Ranjan, A., Fitzgerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. Cell *143*, 725–736.

Luo, Z., Lin, C., Guest, E., Garrett, A., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M., and Shilatifard, A. (2012). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol. Cell. Biol. *32*, 2608–2617.

Lupien, M., Eeckhoute, J., Meyer, C., Wang, Q., Zhang, Y., Li, W., Carroll, J., Liu, X., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958–970.

Mango, S. (2009). The molecular basis of organ formation: insights from the C. elegans foregut. Annu. Rev. Cell Dev. Biol. *25*, 597–628.

Mango, S.E., Lambie, E.J., and Kimble, J. (1994). The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans. Development *120*, 3019–3031.

Mann, R.S., and Carroll, S.B. (2002). Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genet. Dev. *12*, 592–600.

McGhee, J.D., Fukushige, T., Krause, M.W., Minnema, S.E., Goszczynski, B., Gaudet, J., Kohara, Y., Bossinger, O., Zhao, Y., Khattra, J., et al. (2009). ELT-2 is the predominant transcription factor controlling differentiation and function of the C. elegans intestine, from embryo to adult. Dev. Biol. *327*, 551–565.

McGhee, J.D., Sleumer, M.C., Bilenky, M., Wong, K., McKay, S.J., Goszczynski, B., Tian, H., Krich, N.D., Khattra, J., Holt, R.A., et al. (2007). The ELT-2 GATA-factor and the global regulation of transcription in the C. elegans intestine. Dev. Biol. *302*, 627–645.

Missra, A., and Gilmour, D.S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. Proc. Natl. Acad. Sci. USA *107*, 11301–11306.

Muse, G., Gilchrist, D., Nechaev, S., Shah, R., Parker, J., Grissom, S., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Nechaev, S., Fargo, D., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335–338.

Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.R., and Lis, J.T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. Mol. Cell. *13*, 55–65.

O'Brien, R.L., and Born, W. (1991). Heat shock proteins as antigens for gamma delta T cells. Semin. Immunol. *3*, 81–87.

Ohler, U., Liao, G.C., Niemann, H., and Rubin, G.M. (2002). Computational analysis of core promoters in the Drosophila genome. Genome Biol. *3*. doi:10.1186/gb-2002-3-12-research0087

Okkema, P., Ha, E., Haun, C., Chen, W., and Fire, A. (1997). The Caenorhabditis elegans NK-2 homeobox gene ceh-22 activates pharyngeal muscle gene expression in combination with pha-1 and is required for normal pharyngeal development. Development *124*, 3965–3973.

Okkema, P.G., and Fire, A. (1994). The Caenorhabditis elegans NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. Development *120*, 2175–2186.

Priess, J.R., Schnabel, H., and Schnabel, R. (1987). The glp-1 locus and cellular interactions in early C. elegans embryos. Cell *51*, 601–611.

Puri, P.L., and Sartorelli, V. (2000). Regulation of muscle regulatory factors by DNA-binding, interacting proteins, and post-transcriptional modifications. J. Cell Physiol. *185*, 155–173.

Raharjo, I., and Gaudet, J. (2007). Gland-specific expression of C. elegans hlh-6 requires the combinatorial action of three distinct promoter elements. Dev. Biol. *302*, 295–308.

Raharjo, W.H., Logan, B.C., Wen, S., Kalb, J.M., and Gaudet, J. (2010). In vitro and in vivo characterization of Caenorhabditis elegans PHA-4/FoxA response elements. Dev. Dyn. *239*, 2219–2232.

Rahl, P., Lin, C., Seila, A., Flynn, R., McCuine, S., Burge, C., Sharp, P., and Young, R. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Sartorelli, V., Webster, K.A., and Kedes, L. (1990). Muscle-specific expression of the cardiac alpha-actin gene requires MyoD1, CArG-box binding factor, and Sp1. Genes Dev. *4*, 1811–1822.

Simone, C., Stiegler, P., Bagella, L., Pucci, B., Bellan, C., De Falco, G., De Luca, A., Guanti, G., Puri, P.L., and Giordano, A. (2002). Activation of MyoD-dependent transcription by cdk9/cyclin T2. Oncogene *21*, 4137–4148.

Smith, P.A., and Mango, S.E. (2007). Role of T-box gene tbx-2 for anterior foregut muscle development in C. elegans. Dev. Biol. *302*, 25–39.

Sommermann, E.M., Strohmaier, K.R., Maduro, M.F., and Rothman, J.H. (2010). Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and -7 mediates the specification to differentiation transition. Dev. Biol. *347*, 154–166.

Sun, K., Coic, E., Zhou, Z., Durrens, P., and Haber, J.E. (2002). Saccharomyces forkhead protein Fkh1 regulates donor preference during mating-type switching through the recombination enhancer. Genes Dev. *16*, 2085–2096.

Takahashi, H., Parmely, T., Sato, S., Tomomori-Sato, C., Banks, C., Kong, S., Szutorisz, H., Swanson, S., Martin-Brown, S., Washburn, M.*,* et al. (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92–104.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science *242*, 405–411.

Updike, D.L., and Mango, S.E. (2006). Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA**.** PLoS Genet. 2, e161.

Updike, D.L., and Mango, S.E. (2007). Genetic suppressors of Caenorhabditis elegans pha-4/FoxA identify the predicted AAA helicase ruvb-1/RuvB. Genetics *177*, 819–833.

Vilimas, T., Abraham, A., and Okkema, P.G. (2004). An early pharyngeal muscle enhancer from the Caenorhabditis elegans ceh-22 gene is targeted by the Forkhead factor PHA-4. Dev. Biol. *266*, 388–398.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F.*,* et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. *12*, 343–356.

Welboren, W.J., van Driel, M.A., Janssen-Megens, E.M., van Heeringen, S.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J. *28*, 1418–1428.

Whittle, C.M., McClinic, K.N., Ercan, S., Zhang, X., Green, R.D., Kelly, W.G., and Lieb, J.D. (2008). The genomic distribution and function of histone variant HTZ-1 during C. elegans embryogenesis. PLoS Genet. *4*, e1000187.

Wilkins, R.C., and Lis, J.T. (1997). Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. Nucleic Acids Res. *25*, 3963–3968.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. Biochim. Biophys. Acta *1829*, 98–104.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41–51.

Yokoyama, S., Ito, Y., Ueno-Kudoh, H., Shimizu, H., Uchibe, K., Albini, S., Mitsuoka, K., Miyaki, S., Kiso, M., Nagai, A.*,* et al. (2009). A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. Dev. Cell *17*, 836–848.

Zaret, K., and Carroll, J. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512–1516.

Zhong, M., Niu, W., Lu, Z., Sarov, M., Murray, J., Janette, J., Raha, D., Sheaffer, K., Lam, H., Preston, E.*,* et al. (2010). Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1000848

CHAPTER 2

TEMPORAL CONTROL OF PHARYNGEAL GENE EXPRESSION: INTERPLAY BETWEEN THE LEVEL OF PHA-4 AND DNA BINDING AFFINITY

Chapter 2 is a project that will be submitted for publication. Authors are Hui-Ting Hsu and Susan Mango.

2.1 Abstract

The formation of organs is often controlled by selector genes, which encode a special class of transcription factors. In *C. elegans*, the forkhead transcription factor PHA-4/FoxA plays an integral role in all aspects of foregut/pharynx development by directly activating both early and later expressed pharyngeal genes during embryogenesis. The temporal control of pharyngeal expression is fine-tuned by PHA-4 binding site affinity within the target promoters. Mutation of a high affinity PHA-4 binding site to a lower affinity site delays the onset of target gene expression; conversely, increasing binding site affinity advances gene activation. To address the molecular mechanism of PHA-4 controlled temporal gene expression through binding site affinity, we probed the PHA-4 binding to different targets by the Nuclear Spot Assay and Chromatin Immunoprecipitation. PHA-4 was found to bind to its targets before active transcription, and the level of PHA-4 occupancy was affected by the affinity of binding sites within the target promoters. Changing PHA-4 levels by overexpression or reducing PHA-4 expression during development to affect PHA-4 occupancy at its targets resulted in advanced or delayed pharyngeal gene expression. Our results demonstrate that binding site affinity regulates the level of PHA-4 occupancy at target promoters to modulate temporal gene expression. This study provides insights into understanding how a single selector gene orchestrates transcription with the precision required for organ formation.

2.2 Introduction

Organ formation is a complex developmental process involving early cell fate specification followed by cell differentiation and functional morphogenesis. Organogenesis is driven by successive differential gene expression that is often controlled by organ selector genes, which encode a special class of transcription factors. Organ selector genes are defined based on the absolute requirement of their activity on organ formation; organogenesis is blocked at the earliest specification stage in the absence of selector genes (Mann and Carroll, 2002). Organ selector genes function as transcription factors that establish downstream gene expression profiles needed to generate an organ primordium. Ectopic expression of an organ selector gene induces ectopic expression of organ specific genes and can lead to transformation of cell identify (Fukushige and Krause, 2005; Kiefer et al., 2007; Zhu et al., 1998).

pha-4 in *C. elegans* is a well-characterized organ selector gene, which encodes a forkhead transcription factor orthologous to mammalian FoxA and *Drosophila* forkhead transcription factors (Horner et al., 1998; Kalb et al., 1998). PHA-4/FoxA is required for specification of the foregut in mammals (Kaestner, 2010) and pharynx development in *C. elegans* (Mango et al., 1994). The *C. elegans* pharynx is a pumping and feeding organ composed of multiple cell types, including epithelial, neural, gland, valve, muscle, and structural cells. Embryos lacking *pha-4* fail to form the pharynx and die of starvation soon after hatching (Horner et al., 1998; Kalb et al., 1998; Mango et al., 1994). As might be predicted for a selector gene, PHA-4 is the earliest pharyngeal marker and is

expressed throughout the pharynx during development regardless of cell lineage or the adopted fates of subpharyngeal cell types. Previous microarray analyses identified pharyngeal genes expressed at different stages. The promoters of these genes contain at least one consensus PHA-4 binding site “TRTTKRY” (R = A/G, K = T/G, Y = T/C), and their pharyngeal expression is dependent on PHA-4 binding sites (Gaudet and Mango, 2002; Zhong et al., 2010). This result indicates that PHA-4 is a master regulator of pharyngeal transcription.

How PHA-4 controls pharyngeal gene expression patterns during pharynx development is a central question in understanding the mechanism of temporal gene activation. Previous studies from the Mango lab have shown that PHA-4 binds to different TRTTKRY consensus sequences found at endogenous targets with differential affinity and regulates temporal pharyngeal gene expression partly through its binding affinity to target promoters. Mutation of a high affinity PHA-4 binding site to a lower affinity site within the context of a pharyngeal promoter construct delays the onset of promoter-fused reporter gene expression. Conversely, increasing the binding site affinity advances pharyngeal activation of the reporters (Gaudet and Mango, 2002). These data suggest that binding site affinity influences the timing of pharyngeal onset, within the context of natural promoters.

Binding affinity between transcription factors and their recognition consensus sequences is used in many regulatory contexts. For example, the morphogen Dl (Dorsal) forms a concentration gradient along the dorsoventral axis of the *Drosophila* embryo and activates spatial gene expression. Targets

with high affinity Dl sites are expressed at lateral regions with lower Dl levels, whereas targets containing low affinity Dl sites only express at the ventral mesoderm in response to high Dl levels (Stathopoulos and Levine, 2004). Moreover, binding site affinity also contributes to control the level of gene expression. Studies of the transcription factor Pho4 and phosphate homeostasis in yeast revealed that Pho4 induces low levels of *PHO5* by binding to a low affinity site, whereas high levels of *PHO84* are induced due to its high affinity site bound by Pho4 in intermediate phosphate conditions (Lam et al., 2008; Springer et al., 2003). PHO5 is not highly expressed until phosphate starvation induces substantial Pho4 accumulation in the nucleus, which may saturate available binding sites.

These cases all suggest that binding site affinity is sensitive to the concentration of transcription factors. Similar to the proposed PHA-4 affinity model, we hypothesize that high affinity PHA-4 sites, but not low affinity sites, are bound by PHA-4 at early embryonic stages when PHA-4 levels are low. Low affinity PHA-4 targets are bound and activated until later stages when PHA-4 levels are increased. The promoter occupancy of a given transcription factor is predicted to be different among different affinity sites since the transcriptional outcome has proven to be affected. However, the effect of binding site affinity on the level of transcription factor occupancy has not been carefully examined in vivo due to the lack of a well-controlled system. In this study, I examined pharyngeal expression in response to changes in PHA-4 levels and probed PHA-4 binding to different affinity sites in living embryos. The results indicated that

DNA binding site affinity affects the level of PHA-4 occupancy on target promoters. Understanding the basic biophysical dynamics of transcription factor binding strength to different DNA targets is important to help unravel the molecular mechanisms of transcriptional control, such as the downstream recruitment of macromolecular complexes including RNA Polymerase II and general transcription factors.

2.3 Materials and methods

2.3.1 Strains and growth conditions

Strains were provided by the Caenorhabditis Genetics Center and maintained at 20°C, except for the stated strains. Bristol N2 was used as the wild type strain. **KK822** *par-1(zu310ts)V* was maintained at 24°C and shifted to 15°C at the L4 stage overnight to collect embryos. For the Nuclear Spot Assay (NSA), the following strains were used: **SM1630** *cha-1 (p1182); pxEx274 (cha1 + H2AZpro::lacI::cfp + PHA-4::YFP+M05B5.2-Down + lacO + Sperm DNA),* **SM1636** *cha-1(p1182); pxEx275(cha-1 + H2AZpro::lacI::cfp +PHA-4::YFP + M05B5.2-WT + lacO + Sperm DNA),* **SM1703** *cha-1(p1182); pxEx333(cha-1 + H2AZpro::lacI::cfp + PHA-4::YFP + Ceh-22 WT + lacO + Sperm DNA),* **SM1704** *cha-1(p1182); pxEx334(cha-1 + H2AZpro::lacI::cfp+PHA-4::YFP + Ceh-22 Up + lacO + Sperm DNA).* Transgenic worms generated for the NSA were grown at 24°C. For the *pha-4* heat shock experiment, **SM295** *pxEx(HS::PHA-4 + pax-1::GFP + UL8::lacZ + pRF4 + 1 KB ladder + Herring Sperm DNA)* was used and maintained at 15°C. For the *pha-4* RNAi experiment, **OK29** *culs1[ceh-22::GFP +*

pRF4(rol-6(su1006))] was used. For Chromatin Immunoprecipitation (ChIP), **SM1754** *stIs10389 (pha-4::gfp::3xFLAG)*, **SM2078** *stIs10389 (PHA-4::GFP::3XFLAG),* and *pha-4 (q500) rol-9(sc148)* were used.

2.3.2 Antibody stains

For quantification of PHA-4 levels during embryogenesis, wild type embryos were isolated and fixed for staining as described previously (Kiefer et al. 2007). The primary antibodies used were MAB052 (anti-pan histone) at 1:1000 dilution (Millipore Cat # NG1752060) and anti-PHA-4 Ab at 1:2000 dilution (Kaltenbach et al., 2005). The secondary antibodies were Alexafluor 488 goat anti-rabbit (Alexafluor Cat # a11008) and Alexa Fluor 647 goat anti-mouse at 1:200 dilution. Slow Fade with 4',6-diamidino-2-phenylindole (DAPI) (Molecular Probes) was used as a mounting medium.

2.3.3 Image analysis

For quantification of relative PHA-4 protein levels during embryogenesis, in situ PHA-4 staining of different staged embryos was imaged. Z-stack images covering the whole pharynx were taken at 0.5 um increments. The 3D images of multiple embryos at each embryonic stage were reconstructed and analyzed by Volocity imaging analysis software (PerkinElmer). The measurement protocol was created to first select nuclei (objects) in the embryos based on both DAPI and histone staining. The intensity of PHA-4 staining within pharyngeal nuclei was then measured and the staining signal in nonpharyngeal nuclei was

subtracted as background. The PHA-4 level at each embryonic stage was calculated as follows: (PHA-4 intensity in pharyngeal nuclei / the volume of total measured pharyngeal nuclei) – (background signal in nonpharyngeal nuclei/ the volume of nonpharyngeal nuclei). The PHA-4 staining in terminal-staged *par-1* embryos on each slide was analyzed by the same measurement and used to normalize the staining variation from slide to slide. The ratio of PHA-4 intensity during each stage in wild type embryos to the PHA-4 intensity of *par-1* embryos on the same slides was used to compare the relative PHA-4 levels during embryogenesis. The numbers of endodermal cells (for early embryos) and embryonic morphology (for mid- to late embryos) determined the stages of embryos.

2.3.4 *pha-4* heat shock assay

Two-cell embryos were collected from SM259 worms containing *pha-4* heat shock expression constructs. Two-cell embryos were incubated in the 0.2 mL polymerase chain reaction (PCR) tubes at 20°C using a thermalcycler for 75 minutes (2E), 120 minutes (4E), 3 hours (8E), and 5 hours (16E) prior to heat shock at 33°C for 30 minutes as previously described (Kiefer et al., 2007). After heat shock, embryos were recovered at 20°C to reach the desired stages (8E, 16E, and bean) for analysis. Images were acquired from embryos using a Zeiss confocal microscope (LSM 710).

2.3.5 *pha-4* RNAi assay

RNAi induced by bacterial feeding was performed as previously described (Updike and Mango, 2007). HT115 bacteria expressing dsRNA of mCherry and *pha-4* was grown at 37°C overnight. 10 uL overnight culture was used to inoculate 5 mL lysogeny broth (LB) medium and continued to incubate at 37°C. After 6–8 hours, bacterial growth was measured by absorbance at 600 nm. Both mCherry and *pha-4* bacteria cultures were diluted to the same optical density (OD) *pha-4* RNAi was then diluted to 1/2, 1/4, and 1/8 strength with mCherry RNAi culture. Bacteria were spun down at 2400 rpm for 20 minutes and resuspended in LB medium. Standard nematode growth media (NGM) plates were seeded with 50 uL of bacteria on each plate and allowed to dry overnight. 20–25 L4 OK29 worms were plated on RNAi plates and incubated at 25°C overnight. Embryos on the RNAi plates were collected, fixed and stained with anti-GFP (to recognize *ceh-22*::GFP) and anti-PHA-4 antibodies. Expression was analyzed as described above.

2.3.6 DNA constructs

For the Nuclear Spot Assay, the following constructs were used to generate the extrachromosomal arrays: SEM544 (*M05B5.2* wild type promoter), SEM462 (*M05B5.2* down mutation), SEM542 (*ceh-22* wild type promoter), SEM561 (*ceh-22* up mutation), SEM545 (*ceh-22* promoter with a disrupted PHA-4 binding site), and SEM692 (*Ppha-4*::PHA-4::YFP). Primers C2-67-F and C2-67-B were used to amplify the M05B5.2 wild type promoter from SEM544. Primers

C2-67-B and C2-67-del-3 were used to amplify the M05B5.2 down mutation promoter from SEM462. Primers C22-non-F2 and CEH22-B were used to amplify the *ceh-22* wild type promoter from SEM542. Primers ceh-22-per-F and CEH22-B were used to amplify the *ceh-22* up mutation promoter from SEM561. Primers C22-non-F3 and CEH22-B were used to amplify the *ceh-22* promoter with a disrupted PHA-4 site from SEM545. Primer sequences are listed below: C2-67-F (TTGGTCTAGATTGGCAAACAATCTGAAAGCT), C2-67-B (GGTCGGTACCT AGACTATCTGAATAATTGATAATTG), C2-67-del-3 (CATGTCTCGGAGAGAGG AAGG), C22-non-F2 (CATAATCTATATATTTGTCTTGATGGAAATATTTAAGTA TCCG), CEH22-B (TCCAGGATCCACACGTTGAACATCGGCT), ceh-22-per-F (CATAATCTATATGTTTGCCTTGATGGAAATATTTAAG), and C22-non-F3 (CA TAATCTATATACCGGTCTTGATGGAAATATTTAAGTATCCG)

2.3.7 Nuclear spot assay (NSA)

Transgenic lines for the Nuclear Spot Assay were as follows: *M05B5.2* wild type promoter (SM1702), *M05B5.2* promoter with a mutagenized low affinity PHA-4 binding site (SM1630), *ceh-22* wild type promoter (SM1703), and *ceh-22* promoter with a mutagenized high affinity PHA-4 binding site (SM1637 and SM1704). Nuclear spot assays were set up and performed as previously described (Kiefer et al., 2007; Updike and Mango, 2006), and the live images of embryos were acquired using the Olympus FV1000 Confocal Laser Scanning Microscope as described previously (Fakhouri et al., 2010). The number of endoderm cells and the morphology of embryos were used to determine

embryonic stages. Volocity imaging analysis software (PerkinElmer) was used to perform spatial measurement of array size and intensity of PHA-4::YFP. The measurement protocol was generated to select LacI::CFP (citrine fluorescence protein) (array) areas and PHA-4::YFP (yellow fluorescence protein) positive cells using an intensity threshold. After the separation of touching objects, removal of noise, and exclusion of objects smaller than 0.2 um^2, the intensity of PHA-4::YFP in each pharyngeal cell of different-staged embryos was measured. PHA-4::YFP signal that overlapped with LacI::CFP was measured separately and viewed as "PHA-4::YFP binding to the array." The levels of PHA-4 binding to arrays carrying different affinity sites were calculated based on the ratio of PHA-4::YFP intensity on the array to total PHA-4::YFP intensity in the same nucleus.

2.3.8 Embryo staging and crosslinking for chromatin immunoprecipitation

Starved SM2078 L1 larvae from 4–5 6 cm OP-50 plates were washed and grown in 4 large 10 cm HB101 plates until worms were gravid. The worms were then bleached to collect embryos. Embryos were transferred to a large plate without food to obtain a synchronized L1 larval population. The L1 larvae were then transferred and evenly spread to 20–30 large HB101 plates and grown at 20°C. Worms were harvested and bleached after proximately 56–60 hours incubation until young adult worms have only 1–2 fertilized embryos in the gonads. After bleaching the worms, most of the embryos harvested were 1–4 cell embryos. The embryos were suspended in 1 ml M9 buffer and put on an agar

plate without food for aging to desired developmental stages. These early (1–4 cells) embryos were incubated at 20°C for 3 to 3.5 hours to reach the 8E stage, 5–6 hours to reach the bean stage and 8–10 hours to reach the 2-fold stage. The aged embryos were washed and spun down for harvest. The embryo pellets were flash frozen in liquid nitrogen, followed by thawing on ice for one cycle to crack the eggshell. The embryos were cross-linked by 900 uL 1.5% formaldehyde in M9 for 30 minutes with rotation. 100 uL of 1.25M glycine was added and the pellet incubated for 5 minutes to quench the excess formaldehyde. The embryo pellets were then washed 3 times with cold M9 with protease inhibitors (CalBiochem Protease Inhibitor Cocktail Set 1). Washed pellets were then used immediately for ChIP or stored at -80^{o}C.

2.3.9 Chromatin immunoprecipitation (ChIP)

Approximately 300,000 formaldehyde cross-linked embryos were suspended in 300 uL of lysis buffer (50mM Tris pH 8.0, 10mM EDTA pH 8.0, 0.2% SDS, 1x complete ethylenediaminetetraacetic acid (EDTA) protease inhibitor, 1x PhosStop) and incubated on ice for 30 minutes. Using a QSonica Q700 Sonicator, the samples were sonicated at the following settings: 80% amplitude, 30 seconds on, 30 seconds off and 40 cycles. After sonication, the extracts were spun for 15 minutes at 14,000 RPM at 4°C to remove the debris. The supernatant was transferred to new tubes. Chromatin concentration was measured using a Nanodrop and 15 ug of chromatin was used per ChIP reaction. 15 ug of chromatin was diluted in ChIP dilution buffer (0.01% SDS, to 1.1% Triton

X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8.1, 167mM NaCl) to 1 mL and precleared for 1 hour at 4°C with protein A agarose beads. 10% of ChIP material was saved as input. For each ChIP, 5 uL of GFP trap-M antibody (Chromoteck) was used to pull down PHA-4::GFP. The immunocomplexes were then incubated at 4°C on a rotator for 15–17 hours. After incubation, the beads were washed twice by ChIP dilution buffer, twice by low salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.1, 150mM NaCl), twice by high salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.1, 500mM NaCl), once by LiCl buffer (100mM Tris pH 8.0, 500mM LiCl, 1% deoxycholic acid, 1% NP-40) and three times by TE buffer (10mM Tris pH 8.0, 1mM EDTA pH 8.0) for 5 minutes on the rotator. To elute the bound immunocomplexes, 150 uL of elution buffer (50mM $NaHCO_3$, 140mM NaCl, 1% SDS) were added to each tube and heated at 55°C for an hour with vortexing every 5 to 10 minutes. From this point, the ChIP and input samples were then treated the same way to release the DNA. 1 uL of RNaseA (200 mg/mL) was added, followed by incubation at 37°C for 30 minutes. To reverse the cross-links, 2 uL of proteinase K (10 mg/mL) were added and incubated at 65°C overnight. The released DNA was purified using a Qiaquick PCR purification kit (Qiagen) and eluted twice with 50 uL of elution buffer. For construction of the sequencing library, the DNA was purified and concentrated by using MinElute Reaction Cleanup Kit (Qiagen) and eluted twice with 10 uL of elution buffer.

2.3.10 Quantitative PCR

For measuring the promoter copy numbers of different NSA strains, genomic DNA was isolated from individual strains (Fakhouri et al., 2010). Specific primers targeting *M05B5.2* and *ceh-22* promoters as well as *pha-4::yfp* were used and normalized to an internal control (actin, *act-1)*. The relative copy numbers of promoters and exogenous *pha-4::yfp* were determined using $2^{-\Delta\Delta Ct}$ method.

For quantification of ChIP signals, specific primers (see Table A.1 in the Appendix for primer sequences) against PHA-4 targets and genomic regions were used to detect PHA-4 binding. Each primer set used was calibrated by a standard curve using multiple dilutions of template DNA isolated from cross-linked and sonicated chromatin to quantify the enrichment of binding relative to the input signal.

2.4 Results

2.4.1 PHA-4 expression changes dynamically during embryogenesis

The model of affinity suggests that the amount of PHA-4 in the pharyngeal nuclei would impact the pharyngeal gene expression. PHA-4/FoxA is activated at early embryonic stages and maintained throughout the whole lifespan, as determined by GFP-based transcriptional and translational reporters (Azzaria et al., 1996; Horner et al., 1998). However, the dynamic change in PHA-4 expression level during embryogenesis was not well studied until a recent *pha-4*

mRNA expression profile was reported (Levin et al., 2012). The *pha-4* mRNA expression was measured in whole embryos in which *pha-4* expresses in the pharynx, intestine, and rectum. To determine the changes of PHA-4 protein levels specifically in the pharynx during embryogenesis, we employed a sensitive approach to quantify endogenous PHA-4 protein by PHA-4 antibody staining in embryos. The PHA-4 protein level was measured by counting pixels in a region of interest for the pharynx. The protein concentration of PHA-4 at each embryonic stage was calculated based on the total pixels in a measured area divided by the volume of the area. The average protein concentration of PHA-4 was then subtracted by the background staining signal in nonpharyngeal nuclei measured and calculated using the same formula mentioned above. The *par-1* mutant embryos that contain excess pharyngeal cells were mixed with wild type embryos as an on-slide control for antibody staining. The relative PHA-4 protein concentration at each embryonic stage was determined by normalization to PHA-4 levels in *par-1* (Figure 2.1A) (Kaltenbach et al., 2005; Updike and Mango, 2007). PHA-4 protein was first detectable at the 2-endodermal cell (2E) stage, around 60–100 minutes after the first cleavage when embryos contain total ~28 cells). Quantified by protein concentration, PHA-4 expression increased 10-fold during the first half of embryogenesis. PHA-4 expression reached a plateau at the bean stage at which the pharyngeal primordium is formed. PHA-4 expression was maintained during mid-embryogenesis and decreased slightly at later embryonic stages (Figure 2.1B). This trend of dynamic PHA-4 expression during embryogenesis is similar to *pha-4* mRNA profiles (Levin et al., 2012) but

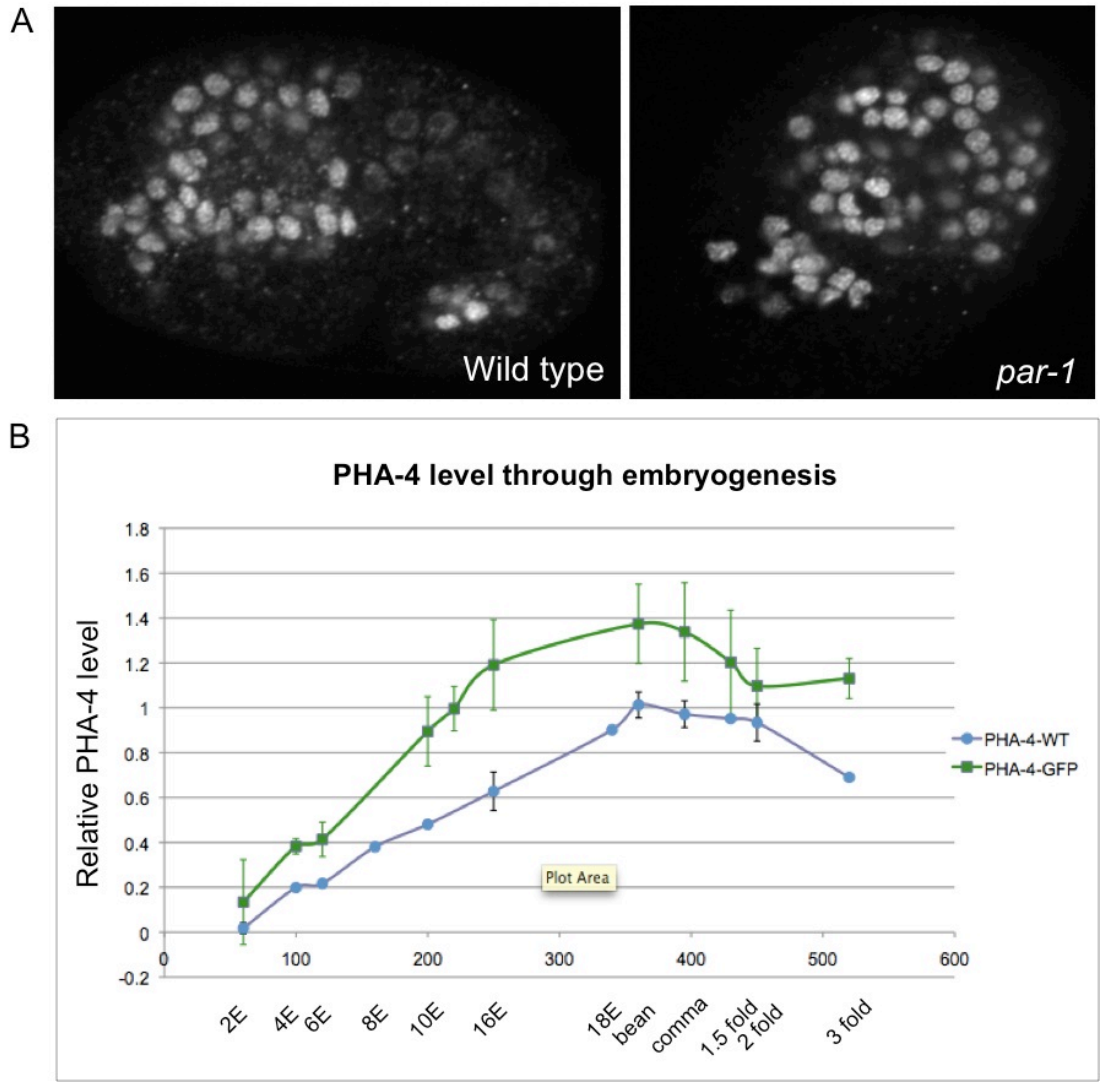


Figure 2.1

PHA-4 expression changes dynamically during embryogenesis. (A) PHA-4 staining patterns in wild type and *par-1* mutant embryos. PHA-4 expression level at each embryonic stage was quantified by measuring the signal intensity of PHA-4 within the pharyngeal cells and subtracting the background staining in nonpharyngeal cells. The ratio of PHA-4 level at each individual stage to PHA-4 intensity in *par-1* mutant embryos on the same slide was used as an internal control to compare staining. (B) The changing of relative PHA-4 level during embryogenesis in wild type and PHA-4::GFP strain. PHA-4 level in the pharyngeal cells increases during the first half of embryogenesis and decreases slightly during the later embryonic stages. PHA-4::GFP strain was the same strain used in modENCODE studies (Zhong et al. 2010) (SM1754; see materials and methods) Timing (minutes after first cleavage) and corresponding embryonic stages at 20°C: 2E (~60-100 min), 4E (~after 100 min), 8E (~180 min), pre-bean (16E) (~280 min), bean (~300-360min), comma (~400 min), 1.5 fold (~440 min), 2 fold (~480 min), and 3 fold (~540 min).

consistently delayed. The delay of timing may reflect the time required for protein synthesis.

Using the same strategy of PHA-4 staining to quantify total PHA-4 levels of the integrated PHA-4::GFP strain (SM1754) utilized for the modENCODE project (Zhong et al., 2010), we determined that the PHA-4 level of PHA-4::GFP strain is 2–3-fold higher than wild type (Figure 2.1B, data shown in green line). This PHA-4::GFP strain (SM1754) is in a wild type genetic background which also expresses endogenous PHA-4. This result suggests that a 1–2-fold increase of PHA-4 protein level is expressed by extra copies of integrated *pha-4::gfp*. In later studies, we generated a *pha-4* rescued strain (SM2078; see materials and methods) by crossing the PHA-4::GFP strain to a *pha-4* mutant (*pha-4(q500);* (Mango et al., 1994) in which PHA-4::GFP is the only source of functional PHA-4 and the total PHA-4 level is similar to wild type.

2.4.2 Adequate PHA-4 is crucial to maintaining precise temporal pharyngeal expression

PHA-4 activity is essential for developmental viability, as evidenced by shifting the *pha-4* temperature-sensitive strain to the restrictive temperature at different stages during development. Eliminating PHA-4 activity at early embryonic stages (after the formation of pharyngeal precursors) caused 100% lethality and a complete absence of the pharynx. In addition, abrogating PHA-4 at later stages (before the pharynx fully developed) resulted in a range of pharyngeal defects, such as Pun or a stuffed pharynx (defects in pharyngeal

pumping) (Gaudet and Mango, 2002; Kiefer et al., 2007). These results indicate that maintaining PHA-4 levels is required for pharyngeal gene expression at both early and late embryonic stages. We first asked if endogenous pharyngeal gene expression is affected in the PHA-4::GFP strain (SM1754), which contains 2–3-fold more PHA-4 protein compared to wild type. Two endogenous pharyngeal muscle targets, a pharyngeal muscle epitope 3NB12 (Priess and Thomson, 1987) and pharyngeal muscle myosin, were examined by antibody staining in SM1754 and wild type embryos. The antibody staining showed that the onset of two pharyngeal muscle targets is normally regulated as in wild type (data not shown). This result implies that temporal pharyngeal expression was maintained properly within a 2–3-fold increase in PHA-4 levels. To understand the pharyngeal defects caused by abnormality of PHA-4 activity, we examined the patterns of pharyngeal gene expression by changing PHA-4 levels in a wider range than 2–3-fold during development. Ectopic PHA-4 expression was placed under control of the heat shock promoter to induce PHA-4 overexpression at early embryonic stages, specifically at 2E, 4E, 8E, and 16E. PHA-4 expression driven by a heat shock promoter was activated ubiquitously in both pharyngeal and nonpharyngeal cells. The outcome of pharyngeal expression was detected by a *Ppax-1*::GFP reporter (a direct PHA-4 target) (Figure 2.2A). In wild type and no heat shock control embryos, *Ppax-1*::GFP was activated from the bean stage (Figure 2.2B). Heatshock *pha-4* at 2E, 4E, and 8E, but not 16E induced advanced *Ppax-1*::GFP expression at 16E. In addition, *Ppax-1*::GFP was ectopically expressed in nonpharyngeal cells whereas in control embryos without

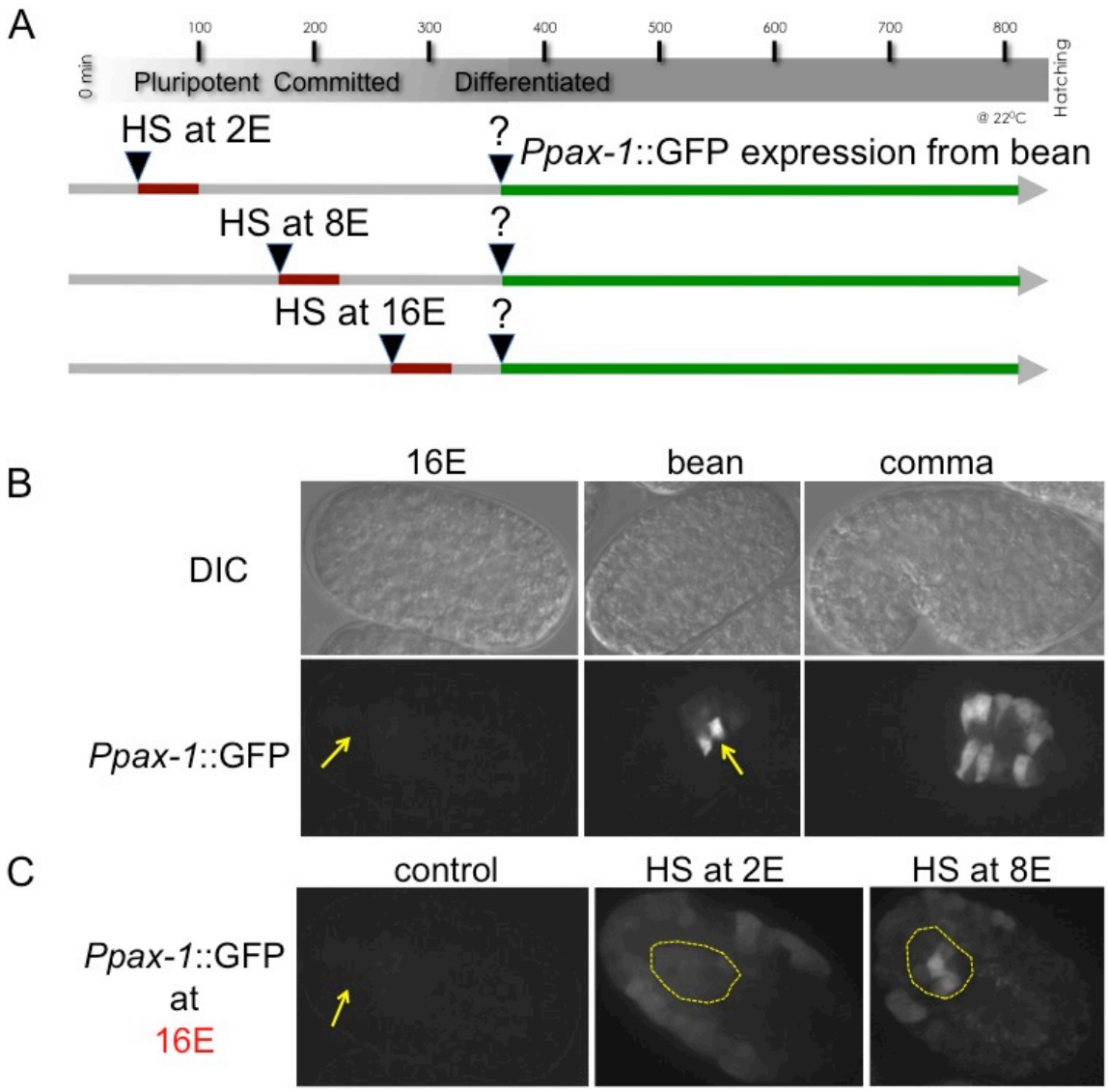


Figure 2.2

Overexpression of PHA-4 induces advanced pharyngeal expression of *pax-1*::GFP. (A) Experimental design: heatshocked *pha-4* at 2E, 8E, and 16E for 30 min and recovered to desired stages. GFP expression was detected at 16E or bean stages. *Ppax-1*::GFP normally turns on at bean stage. (B) The patterns of wild type *Ppax-1*::GFP expression at 16E, bean, and comma stages. *Ppax-*1::GFP was first detectable at bean stage in pharyngeal marginal cells. (C) The patterns of *Ppax-1*::GFP expression at 16E under different conditions. No *Ppax-*1::GFP expression was detected in control embryos without heat shock at 16E. Overexpression of *pha-4* by heat shock at 2E and 8E induced advanced and ectopic *Ppax-1*::GFP expression at 16E. Arrows and circled dash line indicate pharyngeal cells.

heat-shock, *Ppax-1*::GFP was expressed specifically in pharyngeal marginal cells (Figure 2.2B and Figure 2.2C). When embryos were heatshocked at the 2E stage (n = 19), a larger portion of nonpharyngeal cells responded to PHA-4 activity and ectopically expressed *Ppax-1*::GFP. In comparison, heat-shock at 8E (n = 22) induced early onset of *pax-1*::GFP expression at 16E but resulted in fewer nonpharyngeal cells exhibiting ectopic expression. Our data showed that increasing PHA-4 levels during early embryogenesis led to advanced activation of PHA-4 targets.

To test the converse, expression of the pharyngeal muscle target, *Pceh-22*::GFP, was examined under reduced levels of *pha-4* by RNAi. *pha-4* RNAi was diluted to 1/4 and 1/8 strength in order to score enough number of F1 embryos, and *Pceh-22*::GFP was scored using live imaging. The mCherry RNAi was used as a control that did not affect GFP expression. In the mCherry RNAi control embryos, *Pceh-22*::GFP was activated in a subset of pharyngeal muscle cells from the bean stage, similarly to its activation in wild type embryos. In embryos with *pha-4* RNAi, a proportion of bean embryos not expressing *ceh-22*::GFP were observed, while all later embryos (later than comma stage) expressed *Pceh-22*::GFP. The proportion of bean embryos that did not express *Pceh-22*::GFP was variable and likely due to the strength of *pha-4* RNAi. This result from live imaging suggests a delay in *Pceh-22*::GFP expression when PHA-4 levels were reduced by RNAi. To verify that the delayed *Pceh-22*::GFP was due to decreased *pha-4* activity, PHA-4 levels were quantified in bean stage *pha-4* RNAi-treated embryos with or without detectable *Pceh-22*::GFP expression by

antibody staining of PHA-4 and GFP (Figure 2.3). Surprisingly, we found a 10-fold decrease in PHA-4 for delayed *Pceh-22*::GFP bean embryos in comparison to bean embryos that were normally activated (Figure 2.3B). This result suggests that a low amount of PHA-4 (above approximately 10% of wild type PHA-4) is sufficient to maintain the temporal expression of *Pceh-22*::GFP.

2.4.3 PHA-4 binds to its targets long before activation of transcription

Pharyngeal gene expression depends on PHA-4 activity and the PHA-4 binding sequences found in target promoters. PHA-4 can activate pharyngeal gene expression at different developmental stages (Gaudet and Mango, 2002). We speculated that PHA-4 might bind to different targets at different times. For example, early expressed pharyngeal genes are bound by PHA-4 earlier than pharyngeal genes expressed at later embryonic stages. To understand how PHA-4 regulates temporal pharyngeal gene expression through its interaction with target DNA, we employed two different methods to study PHA-4 binding to pharyngeal promoters. The first approach is through Nuclear Spot Assay (NSA), which assays the binding of PHA-4::YFP and LacI::CFP to extrachromosomal arrays carrying multiple copies of a pharyngeal promoter of interest and the Lac operator (Figure 2.4A) (Kiefer et al., 2007; Fakhouri et al., 2010). The binding of LacI::CFP to LacO sequences on the array revealed the location and the size of the extrachromosomal array within the nucleus. If YFP and CFP signals colocalize, then we can conclude that PHA-4::YFP associates with the

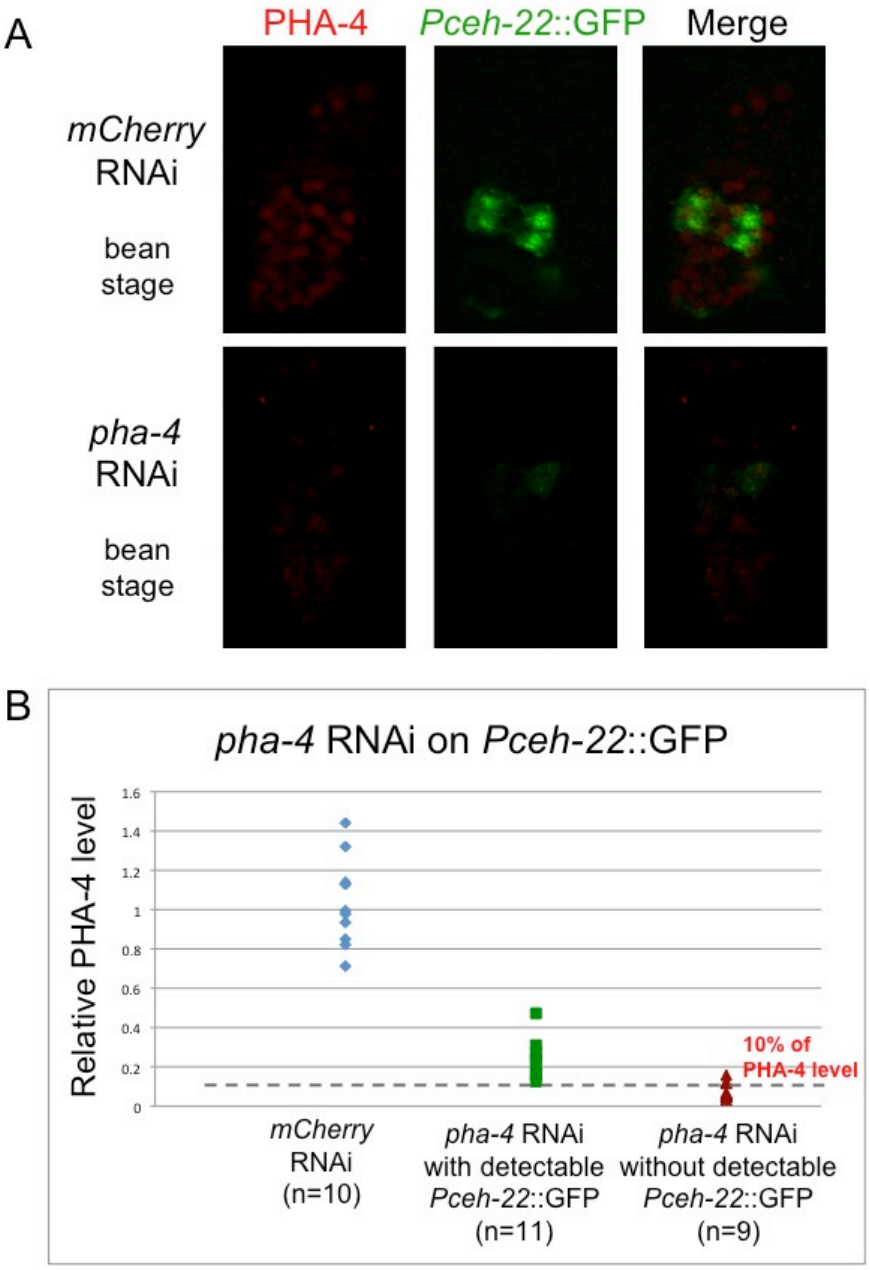


Figure 2.3

Reducing PHA-4 levels results in delayed pharyngeal expression of *Pceh-22*::GFP. (A) Representative images of PHA-4 and GFP staining of bean embryos that were treated with mCherry or *pha-4* RNAi. (B) Quantification of relative PHA-4 level in bean embryos treated with RNAi. Bean embryos treated with *pha-4* RNAi were separated depending on whether or not they had *Pceh-22*::GFP expression. A threshold of 10% of PHA-4 level in *pha-4* RNAi-treated bean embryos distinguished the expression of *Pceh-22*::GFP expression.

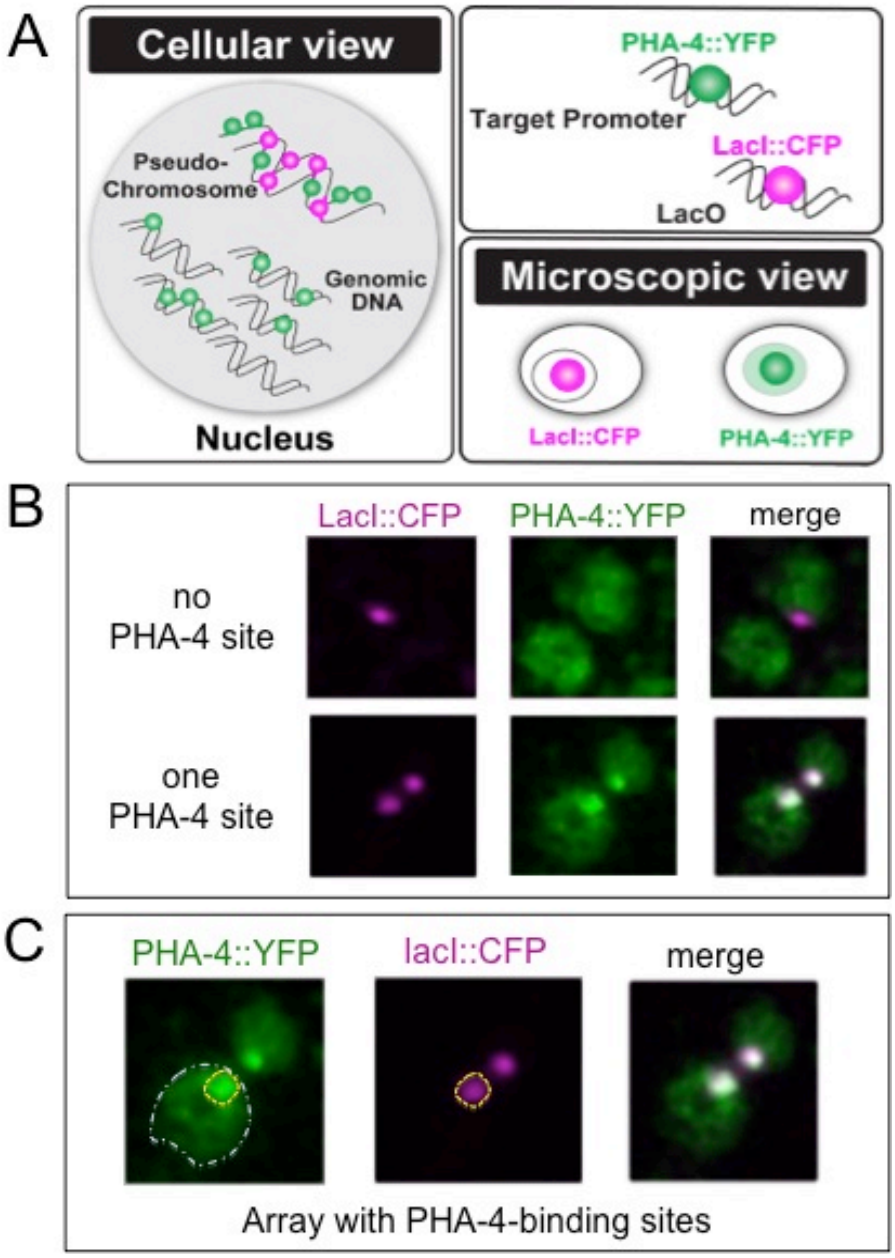


Figure 2.4

Nuclear spot assay to visualize PHA-4 binding to its target DNA. (A) Experimental design of NSA. LacI::CFP (magenta) bound LacO fragments on pseudochromosomes. PHA-4::YFP (green) bound to target promoters on the arrays and also within the genome. (Adapted from Fakhouri et al., 2010) (B) PHA-4::YFP was enriched on the arrays carrying target promoters with wild type PHA-4 binding sites but not on arrays with disrupted sites. Magenta: LacI::CFP; green: PHA-4::YFP; white: colocalization of two signals. (C) The amount of PHA-4 binding to DNA among different affinity sites during development was quantified using Volocity. LacI::CFP intensity was measured to track array size, which reflected the compactness of the pseudochromosome.

pharyngeal promoter on the array. Previous work showed that PHA-4 association with an extrachromosomal array containing a pharyngeal promoter is solely dependent on the availability of PHA-4 binding sites. PHA-4::YFP was exclusive or exhibited a basal intensity with random nuclear distribution at the array carrying a promoter with a disrupted PHA-4 binding site (Figure 2.4B) (Fakhouri et al*.,* 2010). Through live imaging, NSA offers the advantage to study the dynamics of PHA-4 binding to target promoters in the pharyngeal nuclei. Therefore, NSA provides temporal and cell-type specific resolution for studying the binding of transcription factors to target DNA.

NSA allows us to study PHA-4 binding to arrays containing PHA-4 targets but not endogenous loci. To examine PHA-4 binding to endogenous targets, we performed chromatin immunoprecipitation (ChIP), which directly precipitates the transcription factor bound-DNA to determine the endogenous targets of studied factors. Due to the lack of antibody to precipitate PHA-4 directly, we used an alternative approach by generating a transgenic strain which expresses a double-tagged PHA-4::GFP::3xFLAG construct. This PHA-4::GFP::3xFLAG construct rescued the *pha-4* mutant phenotype; hence, we used this rescued strain (SM2078, see materials and methods) to perform ChIP using antibodies recognizing GFP or FLAG. PHA-4::GFP::3xFLAG is the only source of functional PHA-4 in the rescued strain and the level of PHA-4::GFP is similar to wild type PHA-4 levels. This strategy provided higher sensitivity for investigating the level of PHA-4 binding to various targets since there was no competition between endogenous PHA-4 and exogenous PHA-4::GFP::3xFLAG binding to DNA.

By NSA, we examined PHA-4::YFP association with two PHA-4 targets. One was *M05B5.2*, which is broadly expressed in pharyngeal cells from the pre-bean stage; the other target was *ceh-22*, which is activated specifically in a subset of pharyngeal muscles from the bean stage (Figure 2.5 and Figure 2.6) (Gaudet and Mango, 2002). At the 8E stage, before the active transcription of either *M05B5.2* or *ceh-22*, we observed PHA-4::YFP association with the *M05B5.2* and *ceh-22* promoters on the arrays (Figure 2.5B and Figure 2.6B). The levels of PHA-4::YFP association with arrays were variable in individual analyzed pharyngeal nuclei. On average, there was around 10% of total PHA-4::YFP bound to arrays containing the *M05B5.2* or *ceh-22* promoters at the 8E stage. At later embryonic stages, higher levels of PHA-4::YFP association with arrays were detected in more pharyngeal cells. This observation suggests an increase of PHA-4::YFP binding to target promoters during embryogenesis.

To verify the early PHA-4 binding to pharyngeal promoters for endogenous genes, we performed ChIP to detect PHA-4 binding to the cis-regulatory regions of *M05B5.2*, *ceh-22,* and *myo-2* (also a pharyngeal muscle target expressed at late embryonic stages (i.e., ≥ 2-fold)). Similarly, ChIP-qPCR by both GFP and FLAG antibodies showed PHA-4::GFP::3xFLAG association with all three tested pharyngeal genes at the 8E stage, before transcriptional activation (Figure 2.7A and Figure 2.7B). These two approaches revealed that PHA-4 bound to pharyngeal targets earlier before those genes generate mature mRNA.

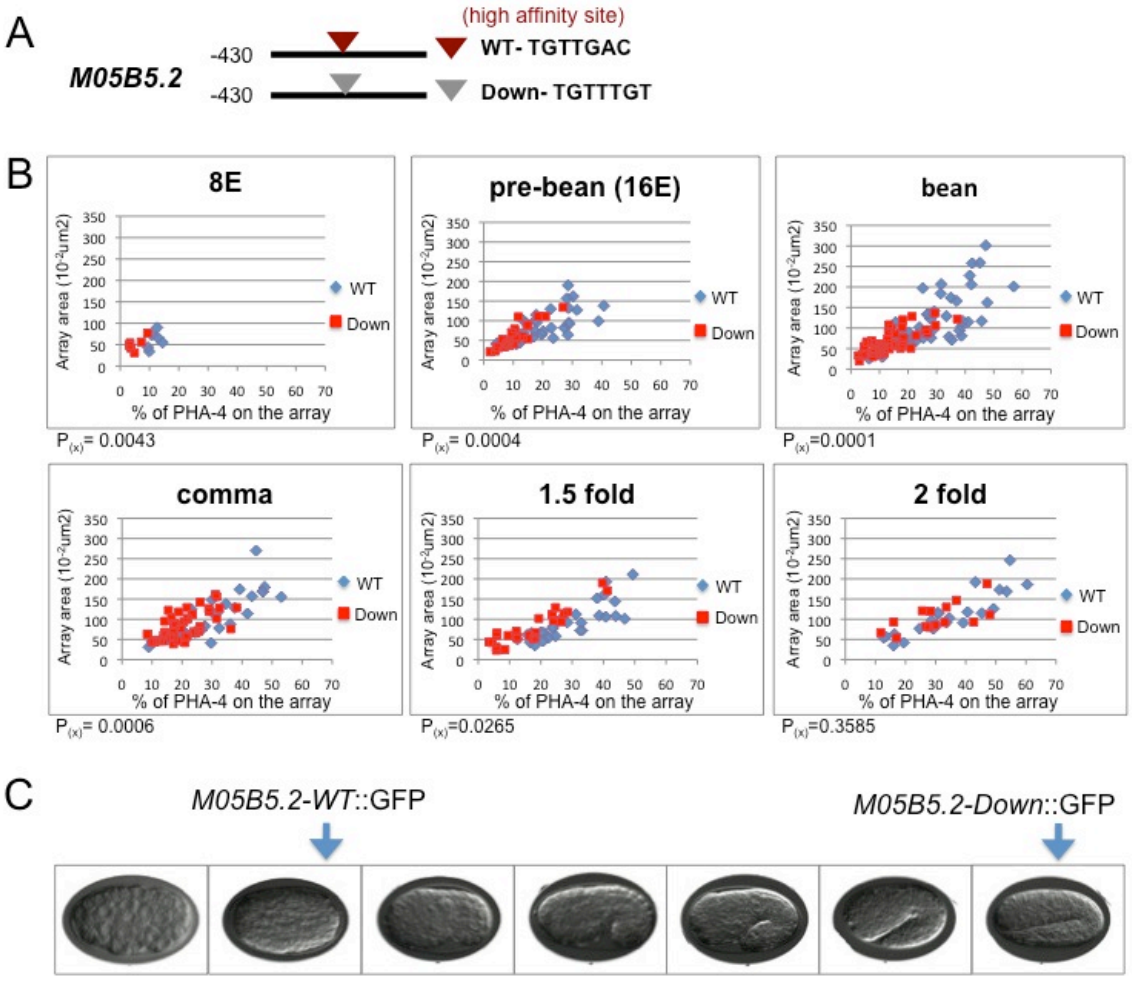


Figure 2.5

Array with lower PHA-4 affinity has reduced PHA-4 binding and is compacted in early stages. (A) Chosen promoter for NSA. PCR amplified 430 bp *M05B5.2* promoters containing a wild type high affinity PHA-4 binding site or a mutated low affinity site (Down mutation) were used for NSA. (B) Measurement of the proportion of PHA-4::YFP binding to the arrays carrying *M05B5.2* promoters with high (WT) or low (Down) affinity PHA-4 sites (X axis) and the size of the arrays (Y axis). NSA was performed by capturing live images of embryos from 8E to 2-fold and analyzed using Volocity imaging software. (C) Temporal expression of *PM05B5.2*::GFP. *M05B5.2-WT* (high affinity) promoter activated GFP reporter at pre-bean whereas M05B5.2-Down (low affinity) promoter delayed GFP expression at 2-fold (Gaudet and Mango, 2002).

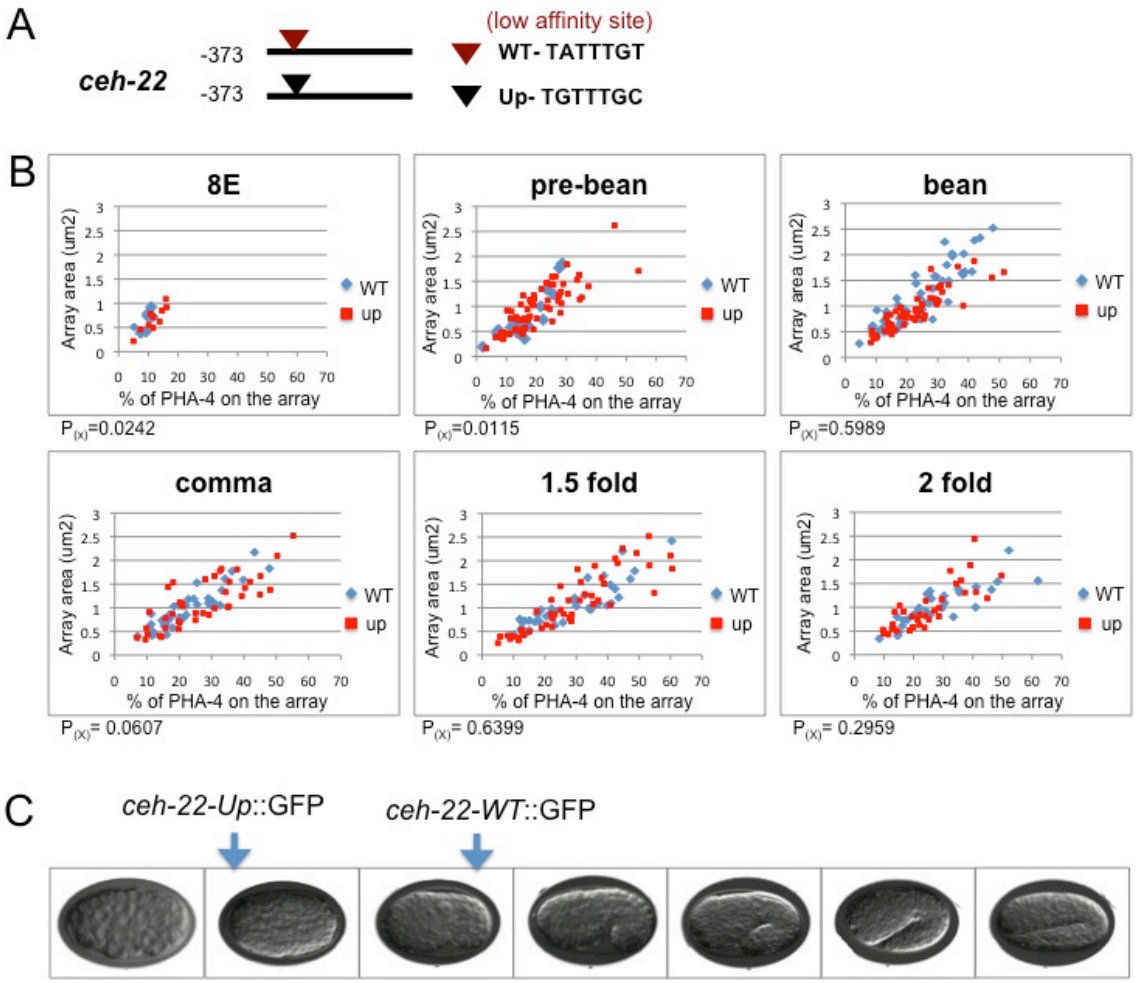


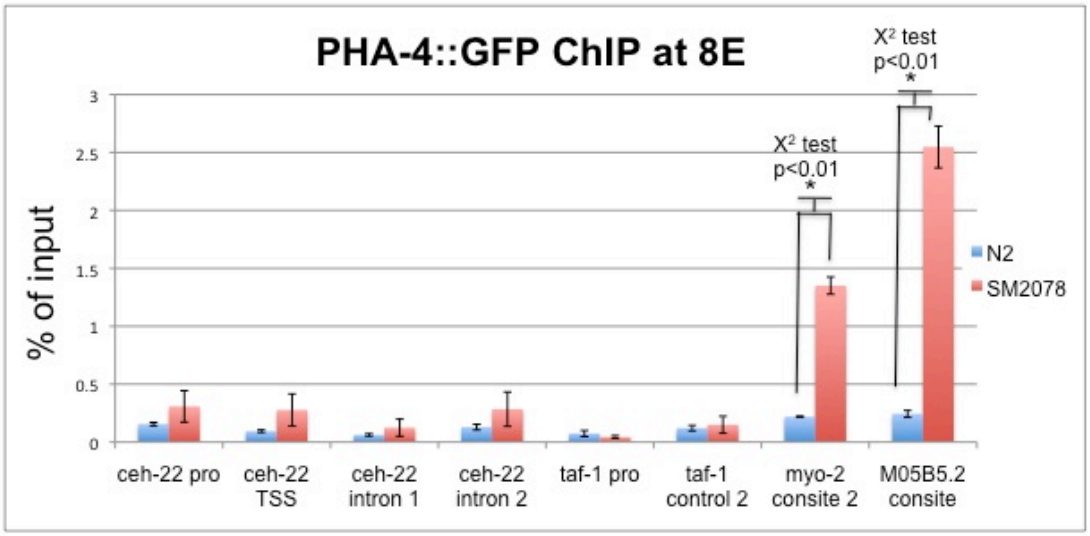
Figure 2.6

Increased binding site affinity enhances PHA-4 binding and is associated with the appearance of a decompacted array at early stages. (A) Chosen promoter for NSA. PCR amplified 373 bp *ceh-22* promoters containing either a wild type low affinity PHA-4 binding site or a mutated high affinity site (Up mutation) were used for NSA. (B) Measurement of the proportion of PHA-4::YFP binding to the arrays carrying the *ceh-22* promoter with low (WT) or high (Up) affinity PHA-4 sites (X axis) and the size of the arrays (Y axis). The NSA was performed by capturing live images of embryos from 8E to 2-fold and data were analyzed using Volocity imaging software. (C) Temporal expression of *Pceh-22*::GFP. The *ceh-22-WT* (low affinity) promoter activated the GFP reporter at bean whereas the *ceh-22-Up* (high affinity) promoter advanced GFP expression at pre-bean (Gaudet and Mango, 2002).

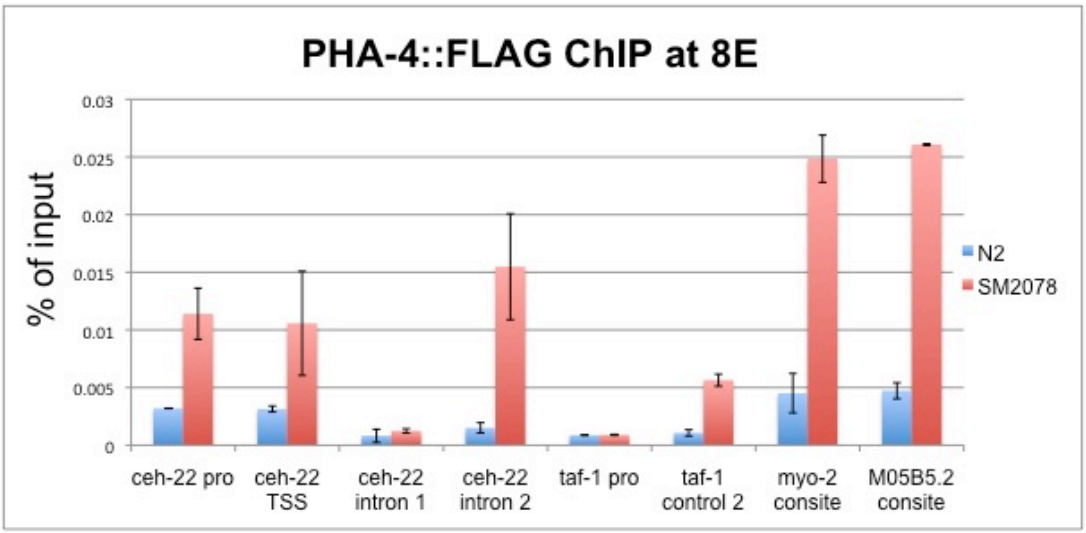
Figure 2.7

PHA-4 binding affinity reflects the level of PHA-4 occupancy at endogenous targets. (A) PHA-4 binding to endogenous targets was detected by PHA-4::GFP ChIP in 8E embryos. (N > 3) (B) PHA-4::FLAG ChIP in 8E embryos. (N = 3) (C) PHA-4::FLAG ChIP in bean embryos. (N = 3) M05B5.2 and myo-2 contain high affinity PHA-4 binding site. ceh-22 pro and ceh-22 TSS contain low affinity PHA-4 binding sites previously identified (Gaudet and Mango 2002). The ceh-22 intron 2 contains a PHA-4 binding element with unknown affinity identified from modENCODE research (Zhong et al. 2010). ceh-22 intron 1, taf-1 pro, and taf-1 control 2 regions contain no PHA-4 binding sites and served as negative controls. N2 without a tagged PHA-4 was used as the wild type strain. The SM2078 strain expresses integrated *pha-4::gfp::flag* in a *pha-4* mutant background.

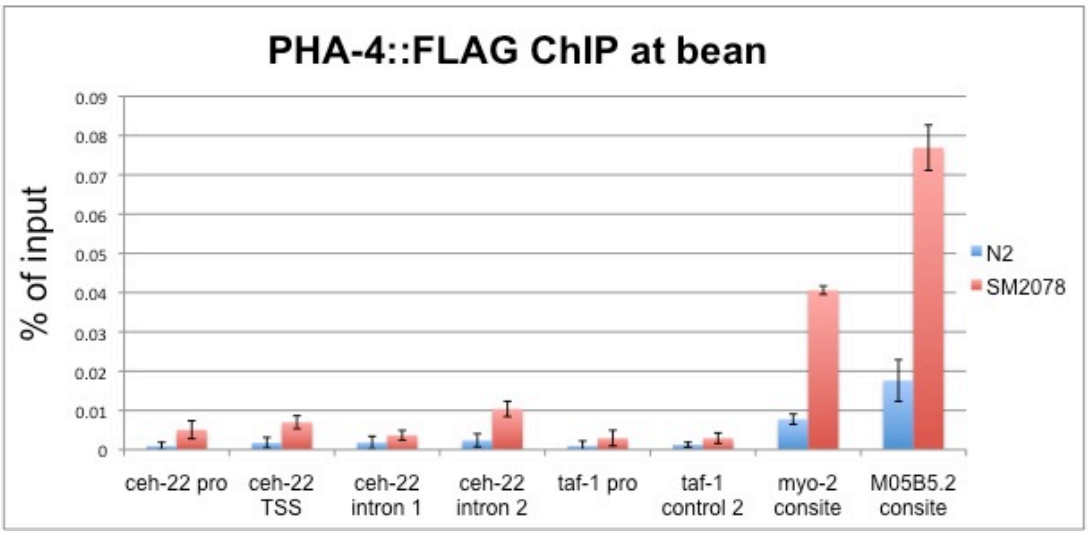
A


PHA-4::GFP ChIP at 8E
X² test p<0.01
*
X² test p<0.01
*
% of input
3
2.5
2
1.5
1
0.5
0
N2
SM2078
ceh-22 pro
ceh-22 TSS
ceh-22 intron 1
ceh-22 intron 2
taf-1 pro
taf-1 control 2
myo-2 consite 2
M05B5.2 consite

B


PHA-4::FLAG ChIP at 8E
% of input
0.03
0.025
0.02
0.015
0.01
0.005
0
N2
SM2078
ceh-22 pro
ceh-22 TSS
ceh-22 intron 1
ceh-22 intron 2
taf-1 pro
taf-1 control 2
myo-2 consite
M05B5.2 consite

C


PHA-4::FLAG ChIP at bean
% of input
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
N2
SM2078
ceh-22 pro
ceh-22 TSS
ceh-22 intron 1
ceh-22 intron 2
taf-1 pro
taf-1 control 2
myo-2 consite
M05B5.2 consite

2.4.4 Binding site affinity affects the level of PHA-4 occupancy at target promoters

Previous studies from our lab indicate that the temporal control of pharyngeal expression is fine-tuned by PHA-4 binding site affinity within the target promoters (Gaudet & Mango, 2002). PHA-4 was able to bind to different target genes before active transcription as determined by NSA and ChIP. We hypothesized that binding site affinity might impact the level of PHA-4 occupancy at target promoters. For instance, the *M05B5.2* promoter and *ceh-22* promoter contain high and low affinity binding sites, respectively, which were previously shown to be important for temporal pharyngeal expression (Gaudet and Mango, 2002). In vitro binding assay indicated that low affinity PHA-4 binding sites (such as PHA-4 site in the *ceh-22* promoter) had 2–5-fold less affinity for PHA-4 than did high affinity sites (Gaudet and Mango, 2002). Within the promoter context, we generated a “down mutation” in *M05B5.2* promoter to lower the PHA-4 binding site affinity and an “up mutation” in the *ceh-22* promoter to increase PHA-4 binding affinity in a range of 2–5-fold (Figure 2.5A and Figure 2.6A). These mutations of PHA-4 affinity sites in the promoter context resulted in shifting the onset of *M05B5.2* and *ceh-22*, as evidenced by the expression of promoter-fused GFP reporters. Specifically, the down mutation to lower PHA-4 binding site affinity within *M05B5.2* promoter caused a delayed GFP expression at the 2-fold stage, which is around 6 hours later in development compared to normal onset at the pre-bean stage (Gaudet and Mango, 2002). To explore the molecular mechanism of PHA-4-controlled temporal gene expression through binding site

affinity, we employed NSA to visualize PHA-4 binding to extrachromosomal arrays carrying target promoters with different PHA-4 affinity sites. We generated strains carrying extrachromosomal arrays with either wild type or a down/up mutation of these PHA-4 target promoters. We matched the strains that bore similar copy numbers of wild type versus mutant PHA-4 target promoters on the arrays in which the available PHA-4 binding sites are equivalent between two compared NSA strains. We further matched the stains expressing similar levels of *pha-4::yfp* so that we could better examine the effect of affinity on PHA-4 binding to target promoters. We measured the level of PHA-4 binding by quantifying the proportion of PHA::YFP colocalized with LacI::CFP (Figure 2.4C). Comparing wild type *M05B5.2* promoter with its down mutation, more PHA-4::YFP was enriched at the array carrying the high affinity PHA-4 binding site, especially at early embryonic stages from the 8E to comma stages ($P < 0.05$) (Figure 2.5B). We also observed slightly increased PHA-4::YFP binding when the wild type low affinity site within *ceh-22* promoter was mutated to a high affinity site. However, PHA-4::YFP binding to the wild type *ceh-22* promoter and its up mutation only showed a difference at the 8E and pre-bean stages ($P < 0.05$) but not at later stages (Figure 2.6B). This NSA result suggests that changing PHA-4 binding site affinity affects the level of PHA-4 association with target promoters. The effect of binding affinity on PHA-4 association is variable between different PHA-4 targets.

Previous studies of PHA-4 and its orthologous FoxA factors indicate that PHA-4/FoxA influences the chromatin environment after binding to its targets

(Cirillo et al., 2002; Kiefer et al., 2007; Updike and Mango, 2006). Using a similar NSA set-up, we have shown that PHA-4 binding to its target promoters results in chromatin decompaction (Fakhouri et al., 2010). To extend this finding, we tested whether binding site affinity plays a role in PHA-4 affected chromatin decompaction by measuring the intensity of LacI::CFP to determine the array size. In the pharyngeal nuclei, we observed a positive correlation of the levels of PHA-4::YFP binding and decompacted array morphology indicated by the increase of array size. When we compared arrays carrying *M05B5.2* promoters with different affinity sites at the same embryonic stages, we found that the arrays with wild type high affinity PHA-4 sites tend to have more PHA-4 binding and are associated with more decompacted array morphology (Figure 2.5B). The down mutation of *M05B5.2* promoter resulted in a less decompacted chromatin at early embryonic stages that might restrain productive transcription and cause delayed onset of gene expression. Overall, our NSA data suggest that changing PHA-4 binding site affinity affects the level of PHA-4 association with target promoters and PHA-4 binding induced chromatin decompaction. The affinity effect is more pronounced at early embryonic stages and variable depending on the different pharyngeal promoters used.

To further investigate the effect of affinity on PHA-4 binding at endogenous targets, we performed ChIP in synchronized early (8E) and later (bean) embryos. From PHA-4 ChIP at the 8E stage, both PHA-4-GFP and PHA-4-FLAG ChIP assays showed a similar trend but different scales in which PHA-4 occupancy was higher at high affinity targets, *M05B5.2* and *myo-2,* compared to

the low affinity target, *ceh-22* (Figure 2.7A and Figure 2.7B). The level of PHA-4 occupancy was significantly different between high versus low affinity PHA-4 targets in the GFP ChIP assay (5–10-fold difference) in comparison with a slightly increased PHA-4 occupancy at high affinity sites (2–2.5-fold) in FLAG precipitation at the 8E stage. Based on the normalization to the input signal, the GFP ChIP assay also showed higher levels of precipitation. This difference suggests that the anti-FLAG antibody was less efficient in precipitating PHA-4::GFP::FLAG. Unfortunately, the anti-GFP antibody deteriorated and lost its specificity, so we were no longer able to use the anti-GFP antibody to perform further ChIP analysis in older embryos. Therefore, we compared whether binding site affinity affected PHA-4 occupancy using the anti-FLAG antibody. When PHA-4 level reaches to a plateau at the bean stage (Figure 2.1B), we still observed more PHA-4 binding at high affinity sites (Figure 2.7C). This result suggests that affinity continuously affects PHA-4 association with different targets at the bean stage. However, the different levels of PHA-4 occupancy at various targets by ChIP might be affected by different cell numbers in which the pharyngeal targets are expressed (see discussion). For this consideration, we compared the PHA-4 occupancy between *ceh-22* and *myo-2*, which are both expressed in a similar number of pharyngeal muscle cells. The FLAG precipitation showed higher PHA-4 occupancy at *myo-2* (high affinity target) in comparison to *ceh-22* (low affinity). This result indicates that binding site affinity affects on the promoter occupancy of transcription factors at endogenous targets.

2.5 Discussion

In this study, we investigated the interplay between the level of a transcription factor PHA-4 and its DNA binding affinity on the level of promoter occupancy and temporal expression of target genes during embryonic development. This study provides two insights into understanding the fundamental mechanisms of transcriptional regulation. First, we used NSA to quantitatively examine the extent to which DNA binding site affinity affected PHA-4 association to different targets within the promoter context. We found that binding site affinity affected the level of PHA-4 occupancy and impacted the temporal pharyngeal expression of promoter-fused reporters (Gaudet and Mango, 2002). Second, the PHA-4 level is critical to control the onset of pharyngeal expression. We tested an affinity model by varying PHA-4 levels and examining the transcriptional output. The onset of pharyngeal gene expression was normal within a 2–3-fold of changing PHA-4 level but was misregulated when PHA-4 level changed dramatically. This observation implies that the transcriptional program is faithfully maintained within a range of transcription factor concentration.

2.5.1 PHA-4 levels and the onset of pharyngeal expression

Genetic analysis of *pha-4* mutant alleles demonstrated that PHA-4 activity is absolutely required for pharynx development (Mango et al., 1994). Reducing PHA-4 activity by shifting *pha-4(ts)* mutant to restrictive temperature at different developmental time points indicates that the level of PHA-4 impacts every aspect

of pharyngeal development and function (Gaudet and Mango, 2002; Kaltenbach et al., 2005). This result suggests that PHA-4 activity is required throughout development and PHA-4-regulated pharyngeal expression is sensitive to PHA-4 dose.

Interestingly, temporal pharyngeal expression was maintained properly within 2–3-fold of increased PHA-4 level when we examined the onset of two endogenous pharyngeal genes expression in the PHA-4::GFP strain. This result is in agreement with the fact that the heterozygous *pha-4/+* animal appears normal (Mango et al., 1994) and suggests that the pharyngeal gene expression is normally regulated. In contrast, *Ppax-1*::GFP was precociously expressed when PHA-4 level was increased dramatically by heat shock in early embryos (Figure 2.2). Similarly, the onset of *Pceh-22*::GFP was delayed when the PHA-4 level was dropped to lower than 10% by *pha-4* RNAi (Figure 2.3). These data support an error tolerance of pharyngeal gene expression in worms, especially the onset of expression, within a range of PHA-4 level fluctuation. Different from worms, heterozygous Foxa/+ mice have various developmental defects, which suggests that a half dose of Foxa is not sufficient to maintain target gene expression (Ang and Rossant, 1994; Weinstein et al., 1994). Maintaining a precise and reproducible transcriptional outcome with a dynamic range of inputs is important for developmental robustness. Perhaps different mechanisms were evolved in different organisms to control transcriptional precision. Utilizing various affinity binding sites to modulate temporal gene expression might provide an advantage to respond to dynamically changing PHA-4 levels during development. This

explanation supports the conclusion that the effect of affinity on PHA-4/FoxA-regulated gene expression is more emphasized in *C. elegans* than other organisms.

In addition to the level of PHA-4, the timing in response to PHA-4 activity is also crucial to set up temporal pharyngeal gene expression. It has been shown that the early embryonic stage between the 2E to 8E stages is a critical time window for cells to respond to developmental regulators, such as PHA-4, and acquire pharyngeal fates (Kiefer et al., 2007). From *pha-4* heat shock experiments, the onset of *Ppax-1*::GFP was advanced and ectopically expressed when *pha-4* was induced between the 2E to 8E stages but not later stages. Based on our observation and previous studies, the data suggest that maintaining proper PHA-4 levels during early embryogenesis specifically during the 2E to 8E stages is important to set up proper temporal pharyngeal expression.

2.5.2 PHA-4 binding affinity and pharyngeal promoter occupancy

To understand how binding site affinity affects transcription factor association with target promoters, we explored PHA-4 occupancy at pharyngeal targets from various angles. First, does changing PHA-4 binding site affinity within the promoter context affect the level of PHA-4 binding? Previously, the research from our lab reported that changing PHA-4 affinity within the promoter's own architecture affects its temporal pharyngeal expression (Gaudet & Mango,

2002). In a similar setting, we examined whether the advanced or delayed pharyngeal expression was caused by increased or decreased PHA-4 occupancy at pharyngeal promoters. Using the NSA, the *M05B5.2* down mutation, which was shown to cause delayed pharyngeal expression (Gaudet and Mango, 2002), indeed showed a decrease in PHA-4 binding compared to a wild type promoter containing a high affinity PHA-4 site. The effect of affinity on PHA-4 binding to *M05B5.2* was most dramatic at early embryonic stages. Surprisingly, the effect of affinity on PHA-4 binding to the *ceh-22* promoter was not significant. Slight differences in PHA-4 occupancy at *ceh-22* wide type (low affinity) and *ceh-22* up mutation (high affinity) were observed at the 8E and pre-bean stages but not later stages.

One possible explanation is the difference of chromatin organization at early versus late embryonic stages. Studies from our lab and others have shown that chromatin is generally open and accessible during early embryogenesis in *C. elegans* (Fakhouri et al., 2010; Meister et al., 2010; Yuzyuk et al., 2009). Therefore, perhaps PHA-4 binding sites are more accessible to PHA-4 binding during early embryogenesis, leading to a more pronounced effect of affinity at early stages. The feature of accessible, decompacted chromatin environment during early embryogenesis supports the early binding of PHA-4 to its targets as we observed with NSA and ChIP.

Based on the effect of affinity observed from NSA, how did PHA-4 occupancy correlate to different affinities among different endogenous pharyngeal targets? PHA-4 ChIP was performed in whole embryos containing

both pharyngeal cells and other cell types. In addition, pharyngeal targets were differentially expressed in either whole pharyngeal cells or a subset of the pharyngeal population. Therefore, it was necessary to take into account the number of pharyngeal cells contributing to the PHA-4 signal at individual pharyngeal promoters. When comparing the PHA-4 occupancy at two pharyngeal muscle targets expressed in a similar number of cells, *ceh-22* and *myo-2*, ChIP showed that the level of PHA-4 occupancy correlated with affinity. The PHA-4 occupancy was 2–5-fold higher at *myo-2* (high affinity site) than at *ceh-22* (low affinity site, specifically in region of ceh-22 TSS detected by ChIP-qPCR, which contains a low affinity site previously identified) at the 8E and bean stages. This result matched the in vitro binding affinity of PHA-4 sites measured previously (Gaudet and Mango, 2002). Accordingly, PHA-4 occupancy at *M05B5.2* that expressed broadly in whole pharyngeal cells showed a 3-fold increase as the number of pharyngeal cells increasing from 8E to bean. Compared to *myo-2* and *ceh-22* that only expressed in pharyngeal muscle cells (37 cells at the bean stage), *M05B5.2* showed the highest PHA-4 occupancy at bean stage (94 cells). This result suggests that both binding site affinity and the cell numbers contributed to the difference of PHA-4 occupancy at various pharyngeal targets detected by ChIP. After carefully examining the differences in cell numbers contributing to the ChIP signal at various PHA-4 targets, we concluded that binding site affinity affected the level of PHA-4 occupancy at endogenous targets.

In summary, we have shown that PHA-4 bound its targets before active transcription. The PHA-4 affinity sites within the target promoters affected the

level of PHA-4 occupancy. The difference in PHA-4 occupancy at pharyngeal promoters further modulated the temporal expression during pharynx development. The proposed affinity model predicts that PHA-4-regulated gene expression is sensitive to PHA-4 concentration as we observed in *pha-4* RNAi and *pha-4* heat shock experiments. Expanding beyond individual pharyngeal gene studies, it will be insightful to examine the genome-wide temporal PHA-4 binding patterns at endogenous loci. The genome-wide study of temporal PHA-4 binding will expand our understanding about binding site affinity involved global gene regulation during organogenesis.

2.6 References

Ang, S.L., and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. Cell *78*, 561–574.

Azzaria, M., Goszczynski, B., Chung, M.A., Kalb, J.M., and McGhee, J.D. (1996). A fork head/HNF-3 homolog expressed in the pharynx and intestine of the Caenorhabditis elegans embryo. Dev. Biol. *178*, 289–303.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Fakhouri, T., Stevenson, J., Chisholm, A., and Mango, S. (2010). Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1001060

Fukushige, T., and Krause, M. (2005). The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early C. elegans embryos. Development *132*, 1795–1805.

Gaudet, J., and Mango, S. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821–825.

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., and Mango, S.E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. Genes Dev. *12*, 1947–1952.

Kaestner, K. (2010). The FoxA factors in organogenesis and differentiation. Curr. Opin. Genet. Dev. *20*, 527–532.

Kalb, J., Lau, K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P., and McGhee, J. (1998). pha-4 is Ce-fkh-1, a fork head/HNF-3alpha, beta, gamma homolog that functions in organogenesis of the C. elegans pharynx. Development *125*, 2171–2180.

Kaltenbach, L.S., Updike, D.L., and Mango, S.E. (2005). Contribution of the amino and carboxyl termini for PHA-4/FoxA function in Caenorhabditis elegans. Dev. Dyn. *234*, 346–354.

Kiefer, J., Smith, P., and Mango, S. (2007). PHA-4/FoxA cooperates with TAM-1/TRIM to regulate cell fate restriction in the C. elegans foregut. Dev. Biol. *303*, 611–624.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246–250.

Levin, M., Hashimshony, T., Wagner, F., and Yanai, I. (2012). Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev. Cell *22*, 1101–1108.

Mango, S.E., Lambie, E.J., and Kimble, J. (1994). The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans. Development *120*, 3019–3031.

Mann, R.S., and Carroll, S.B. (2002). Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genet. Dev. *12*, 592–600.

Meister, P., Towbin, B.D., Pike, B.L., Ponti, A., and Gasser, S.M. (2010). The spatial dynamics of tissue-specific promoters during C. elegans development. Genes Dev. *24*, 766–782.

Priess, J.R., and Thomson, J.N. (1987). Cellular interactions in early C. elegans embryos. Cell *48*, 241–250.

Springer, M., Wykoff, D.D., Miller, N., and O'Shea, E.K. (2003). Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. PLoS Biol. *1*, E28.
doi: 10.1371/journal.pbio.0000028

Stathopoulos, A., and Levine, M. (2004). Whole-genome analysis of Drosophila gastrulation. Curr. Opin. Genet. Dev. *14*, 477–484.

Updike, D., and Mango, S. (2006). Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. PLoS Genet. *2*.
doi: 10.1371/journal.pgen.0020161

Updike, D.L., and Mango, S.E. (2007). Genetic suppressors of Caenorhabditis elegans pha-4/FoxA identify the predicted AAA helicase ruvb-1/RuvB. Genetics *177*, 819–833.

Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., and Darnell, J.E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell *78*, 575–588.

Yuzyuk, T., Fakhouri, T., Kiefer, J., and Mango, S. (2009). The polycomb complex protein mes-2/E(z) promotes the transition from developmental plasticity to differentiation in C. elegans embryos. Dev. Cell *16*, 699–710.

Zhong, M., Niu, W., Lu, Z., Sarov, M., Murray, J., Janette, J., Raha, D., Sheaffer, K., Lam, H., Preston, E.*,* et al. (2010). Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1000848

Zhu, J., Fukushige, T., McGhee, J.D., and Rothman, J.H. (1998). Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes Dev. *12*, 3809–3814.

CHAPTER 3

TEMPORAL POISED RNA POL II DURING C. ELEGANS EMBRYOGENESIS: A POTENTIAL ROLE OF PHA-4/FOXA TO REGULATE POL II

Chapter 3 is a project that will be submitted for publication after additional experiments are completed. Authors are Hui-Ting Hsu and Susan Mango.

3.1 Abstract

Recent studies of genome-wide Pol II occupancy in various organisms and developmental contexts revealed that RNA Pol II is pervasively poised at promoter proximal regions. Therefore, two aspects of gene regulation are suggested: the specific recruitment of Pol II and the control of productive elongation by Pol II at target genes which are regulated by sequence specific transcription factors in the course of development. As studies of poised Pol II in flies and mammals reveal its important role in regulating differential gene expression, Pol II poising in worms is still largely unknown. To determine whether Pol II is poised during early embryogenesis in *C. elegans*, we performed Pol II ChIP-Seq to examine the genome-wide Pol II occupancy in synchronized embryos at early and mid-embryonic stages. Temporal Pol II ChIP-Seq revealed at least four different patterns of Pol II occupancy that reflected the transcriptional status of genes. Moreover, we identified additional genes, including a subset of pharyngeal specific genes that were specifically poised at different embryonic stages, which suggests that Pol II poising is temporally regulated. Interestingly, promoter occupancy of Pol II at poised pharyngeal genes was greatly decreased when PHA-4 was eliminated, indicating that PHA-4 activity is required to load Pol II at a subset of pharyngeal genes. We hypothesize that PHA-4/FoxA functions as a pioneer factor and primes gene activation partly through affecting temporal Pol II occupancy at pharyngeal promoters. These results suggest that other Fox factors, perhaps also other selector genes, might function similarly to control gene expression through regulation of Pol II poising.

3.2 Introduction

The regulation of RNA polymerase II plays a central role in controlling differential gene expression during development, including organogenesis. The specificity of gene expression required for developmental processes is governed by sequence specific transcription factors, such as selector genes that can control a cohort of genes responsible for forming organs or specifying tissue types (Mann and Carroll, 2002). The traditional view of transcriptional control of gene expression involves the recruitment of RNA Pol II to specific promoters. The recent discovery that Pol II pervasively paused at promoter proximal regions in various organisms reveals another layer of transcriptional control (Muse et al., 2007; Rahl et al., 2010; Zeitlinger et al., 2007). These studies suggest two tiers of regulation, specific recruitment of Pol II and control of Pol II productive elongation at target genes. Sequence specific transcription factors could regulate either or both steps to modulate gene expression in response to development.

RNA Pol II is recruited to genes to form a preinitiation complex (PIC), an event that is aided by some transcription factors. For example, some transcription factors recruit histone modifiers and chromatin remodelers resulting in decompaction of chromatin to facilitate recruitment of general transcription factors and RNA Pol II. For example, MyoD binding to the myogenic targets recruits p300 histone actyltransferase and switch/ sucrose nonfementable (SWI/SNF) chromatin-remodeling complexes to create a permissive chromatin environment for PIC assembly (Forcales, 2012; Puri et al., 1997; Simone et al., 2004). Studies with the *Myogenin* promoter have found that MyoD directly

interacts with the cell-type-specific TATA binding protein (TBP) associated factor (TAF) TAF3/TRF3 to assemble the PIC for *Myogenin* activation (Deato et al., 2008). These findings indicate that sequence-specific transcription factors can directly interact with general transcription factors (GTFs) (Fuda et al., 2009). However, the order of the recruitment to build the PIC during transcriptional initiation is still not clear. The establishment of transcriptional initiation might be dependent on the context of the promoters and associated activators rather than on a fixed sequence of steps (Morse, 2007).

On the other hand, sequence-specific transcription factors can also control gene expression through the release of paused Pol II. p-TEFb, composed of cyclin-dependent kinase 9 and cyclin T, is required for the transition from transcriptional initiation to productive elongation (Zhou et al., 2012). Various sequence-specific transcription factors interact with p-TEFb to promote elongation, including NF-κB (Barboric et al., 2001), androgen receptor (Lee et al., 2001), Myc (Eberhardy and Farnham, 2002; Rahl et al., 2010), MEF2 (Nojima et al., 2008), and MyoD (Giacinti et al., 2006). Among these reported transcription factors, MyoD in mammals is a selector gene that regulates target gene expression at both transcriptional initiation and productive elongation steps to control skeletal myogenesis. Besides MyoD, forkhead factors in yeast have been implicated to coordinate pre-mRNA processing through promoting transcriptional elongation (Morillon et al., 2003). The role of selector genes in regulation of transcriptional elongation still waits to be explored.

Although Pol II pausing is implicated as a mechanism to control gene expression, how is paused Pol II established prior to gene activation at specific promoters? Sequence-specific transcription factors likely play a role in this process to provide specificity. GAGA factor binding to the GAGA motif, which is enriched at paused promoters in *Drosophila* embryos (Lee et al., 2008), is thought to recruit chromatin-remodeling complexes and interact with transcription factor II D (TFIID) to maintain an open chromatin environment suitable for establishing paused Pol II (Adkins et al., 2006; Leibovitch et al., 2002). Removing the GAGA motif within a paused promoter did not inhibit gene activation but did eliminate 5' enriched Pol II pausing (Wilkins and Lis, 1997). However, GAGA factor only accounts for about 20% of paused genes in *Drosophila* embryos (Lee et al., 2008), which suggests other factors are required for establishing Pol II pausing at promoters without the GAGA motif. Supporting this idea, no obvious GAGA factor has been found in worms or mammals, despite the existence of poised Polymerase in these organisms (Gilmour, 2009).

Pioneer factors are a potential candidate to establish paused Pol II. Pioneer transcription factors are unique based on their capability to bind target DNA within a compacted chromatin before gene activation (Cirillo et al., 2002; Zaret and Carroll, 2011). In particular, FoxA factors in mammals have been long thought to function as a pioneer factor with the ability to bind the albumin *alb1* enhancer when the chromatin is still compacted, before albumin expression. The binding of FoxA to the *alb1* enhancer in vitro promotes decompaction of chromatin that is sensitive to DNase (Cirillo et al., 2002). Moreover, the binding of

FoxA enables subsequent binding of other factors, such as estrogen receptor (ER), which has been shown in different cell types (Carroll et al., 2005; Lupien et al., 2008; Zhang et al., 2005). Studies from FoxA factors suggest that pioneer factors can function actively to facilitate access of additional factors, such as assembly of the general transcription machinery. Thus, FoxA serves as a good candidate to test if pioneer factors also function to recruit or set up Pol II at specific promoters.

In *Drosophila*, a maternal Zinc finger transcription factor, Zelda, is essential for gene activation during maternal-zygotic transition (Liang et al., 2008). Zelda binds to a specific cis-element, CAGGTAG that is found at promoters and enhancers of many developmental regulators hours before their expression (Liang et al., 2008). Zelda binding is highly correlated with association of many transcription factors at promoters and enhancers genome-wide (Satija and Bradley, 2012), which supports a model that Zelda facilitates the recruitment of other transcription factors to promote zygotic transcription. The early binding of Zelda to target genes in preparation of later expression suggests Zelda is a pioneer factor. In addition, about one-third of Zelda target genes are highly paused in early *Dosophila* embryos (Saunders et al., 2013). To understand how Zelda target genes are regulated, Saunders and colleagues performed GRO-Seq at two time points during early embryogenesis in fly embryos. Their data indicated that Zelda-bound targets were regulated via both recruiting Pol II and releasing paused Pol II, suggesting Zelda does not regulate transcription at the recruitment step, at least not exclusively (Saunders et al., 2013). However, no

one has yet examined whether Zelda is required to recruit Pol II to promoters or to facilitate elongation, using mutant analysis.

The *C. elegans* FoxA homolog, PHA-4 has been implicated to function like a pioneer factor for foregut formation. PHA-4 binds to pharyngeal targets long before the onset of transcription and induces chromatin decompaction partly through recruitment of histone variant H2AZ/HTZ-1 (Fakhouri et al., 2010; Updike and Mango, 2006). Although Pol II ChIP-Seq and GRO-Seq experiments suggested that only 2% or fewer genes exhibited stalled Pol II in mixed-stage embryos, PHA-4 binding clearly overlapped with 85% of Pol II stalled genes in embryos (Kruesi et al., 2013; Zhong et al., 2010). This result reveals a correlation between PHA-4 binding and Pol II poising.

Genome-wide Pol II occupancy in *C. elegans* has been performed in mixed-stage embryos. Due to the heterogeneity of stages, the Pol II signal at each gene is contributed from nuclei with different states of transcriptional activity. This heterogeneity makes it difficult to uncover the patterns of Pol II over time. An ideal experiment would be to analyze Pol II signal within nuclei of a homogeneous population of one cell type at one stage of development. However, the lack of *C. elegans* cell lines and the large quantity of chromatin required for a single ChIP experiment make this goal unattainable. In *Drosophila*, poised Pol II at developmental genes during early embryogenesis was not restricted to one particular cell type (Zeitlinger et al., 2007), suggesting that the ideal experiment may not be essential to characterize Pol II poising in development. However, the dynamics that have been observed in *Drosophila* embryos and *C. elegans* larvae

(Baugh et al., 2009; Gaertner et al., 2012) suggest that staged embryos would be important to track poising during embryogenesis.

To investigate whether Pol II poising is happening during early embryogenesis in *C. elegans*, I collected stage-specific embryos for Pol II ChIP-Seq. Pol II occupancy was analyzed at the 8E stage (an early embryonic stage) and at the bean stage (a midembryonic stage) to examine the Pol II dynamics before and after different gene expression.

PHA-4 activity is absolutely required to control pharyngeal gene expression both early during specification and later during differentiation (Mango, 2009). The features of early binding of PHA-4 to pharyngeal targets and the induction of open chromatin suggest that PHA-4 might recruit Pol II or establish the chromatin configuration to promote Pol II poising before gene activation. I tested this hypothesis by examining Pol II occupancy at pharyngeal targets in wild-type and *pha-4* mutant embryos.

3.3 Materials and methods

3.3.1 Strains and growth condition

Strains were provided by the Caenorhabditis Genetics Center and maintained at 20°C, except for the stated strains. Bristol N2 was used as a wild type strain. SM190 s*mg-1(cc546ts)I;pha-4(zu225)V* and SM568 *smg-1(cc546ts)I; pha-4(q500) rol-9(sc148)V* were maintained at 24°C (permissive temperature) and shifted to 15°C (restrictive temperature) for experimental tests as previously described (Gaudet and Mango, 2002; Kaltenbach et al., 2005).

3.3.2 Embryo staging and crosslinking

To collect stage-specific wild type N2 embryos, starved L1 larvae from 4–5 6 cm OP-50 plates were washed and grown in 250 mL S medium with NA22 as a food source and incubated at 20°C with 200 rpm shaking until worms were gravid. The worms were then bleached to collect embryos. Embryos were transferred to 100 mL S medium without food to obtain a synchronized L1 larval population. The L1 larvae were then transferred to S medium with concentrated NA22 to grow at 20°C at 200 rpm shaking. Worms were harvested and bleached after proximately 56–60 hours incubation, when young adult worms have only 1–2 fertilized embryos in the gonads. After bleaching the worms, most of the embryos harvested contained 1–4 cells embryos. The embryos were suspended in 1mL M9 buffer and put on a NGM plate without food to age to the desired developmental stages. The 1–4 cells embryos were incubated at 20°C for 3 to 3.5 hours to reach the 8E stage, 5–6 hours to the bean stage and 8–10 hours to the 2-fold stage. The aged embryos were washed to harvest and spun down. The embryo pellets were flash frozen in liquid nitrogen following thawing on ice for one cycle to crack the eggshell. The embryos were cross-linked by adding 900 uL 1.5% formaldehyde in M9 and incubating for 30 minutes with rotation. 100 uL of 1.25M glycine was added and incubated for 5 minutes to quench the excess formaldehyde. The embryo pellets were then washed 3 times with cold M9 with protease inhibitors (CalBiochem Protease Inhibitor Cocktail Set 1). After the wash, the embryo pellets were ready for ChIP.

For collecting *pha-4(ts)* mutant embryos, SM190 were grown on 10 large 10 cm plates with HB101 as a food source and maintained at 24°C for two and half days until worms were gravid. The worms were then bleached to collect embryos. Embryos were transferred to a large 10 cm NGM plate without food to obtain a synchronized L1 larval population. The L1 larvae were then transferred to 10 cm plates seeded with HB101 bacteria at a density of 50,000 L1 per large plate. After growing at 24°C for around 40 hours, the majority of worms were L3 to L4 stage. At this point, *pha-4(ts)* worms were shifted to 15°C and incubated overnight. Worms were harvested and bleached when young adult worms had 1–4 fertilized embryos in the gonads; most of the embryos harvested contained 1–8 cells embryos. The embryos were suspended in 1 mL M9 buffer and put on an NGM plate without food and aged to the desired developmental stages. 1–8 cells embryos were incubated at 15°C for 5–8 hours to achieve a population containing embryos from the 8E to bean stages. The aged embryos were prepped as wild type embryos for ChIP.

3.3.3 Chromatin immunoprecipitation

Approximately 300,000 embryos were suspended in 300 uL of lysis buffer (50mM Tris pH 8.0, 10mM EDTA pH 8.0, 0.2% SDS, 1x complete EDTA protease inhibitor, 1x PhosStop) and incubated on ice for 30 minutes. Using a QSonica Q700 Sonicator, the samples were sonicated at the following settings: 80% amplitude, 30 seconds on, 30 seconds off, and 20 cycles. After sonication, the extract was spun down for 15 minutes at 14,000 rpm at 4°C to remove the

debris. The supernatant was transferred to new tubes. Chromatin concentration was measured by Nanodrop and 50ug of chromatin was used per ChIP reaction. 50 ug of chromatin was diluted in ChIP dilution buffer (0.01% SDS, to 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8.1, 167mM NaCl) to 1 mL and precleared for 1 hour at 4°C by protein A agarose beads. 10% of ChIP material was saved as input. A commercially available antibody, 4H8 (Covance), was used to detect RNA polymerase II. 2 ug of 4H8 antibody was conjugated to 25 uL of protein G magnetic beads for at least 6 hours at 4°C prior to ChIP. The immunocomplexes were then incubated at 4°C on a rotator for 15–17 hours. After incubation, beads were washed twice with ChIP dilution buffer, twice with low salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.1, 150mM NaCl), twice with high salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.1, 500mM NaCl), once with LiCl buffer (100mM Tris pH 8.0, 500mM LiCl, 1% deoxycholic acid, 1% NP-40), and three times with TE buffer (10mM Tris pH 8.0, 1mM EDTA pH 8.0) for 5 minutes on the rotator. 1 mL of buffer was used in every wash and briefly vortexed for 10 seconds before 5 minutes incubation on the rotator. To elute the bound immunocomplexes, 150 uL of elution buffer (50mM $NaHCO_3$, 140mM NaCl, 1% SDS) was added to each tube and heated at 55°C for an hour with vortexing every 5 to 10 minutes. From this point, the ChIP input samples were then treated the same to release DNA. Brifely, 1uL of RNaseA (200 mg/mL) was added to each tube followed by incubation at 37°C for 30 minutes. To reverse cross-links, 2 uL of protease K (10 mg/mL) were added and incubated at 65°C overnight. The released DNA was

purified using a Qiaquick PCR purification kit (Qiagen) and eluted twice with 50 uL of elution buffer. For sequencing library construction, the DNA was purified and concentrated using a MinElute Reaction Cleanup Kit (Qiagen) and eluted twice with 10 uL of elution buffer. Both input DNA and ChIP DNA was loaded in a 2% low melting agarose gel to select fragment sizes lower than 500 bp before performing quantitative PCR.

3.3.4 Quantitative PCR

For quantification of ChIP signals and gene expression, specific primers against genomic regions of PHA-4 targets, positive control (*eft-3*) and negative control (*SRW-99 and SRW-96*) were used to detect Pol II occupancy and RNA expression. Each primer set (see Table A.1 for primer sequences) was calibrated by a standard curve using multiple dilution of template DNA isolated from cross-linked and sonicated chromatin to quantify the enrichment of binding relative to the input signal. For our experiment, KAYA SYBR Fast Universal QPCR Kit was used (KAPA Biosystems Cat# KK4602).

3.3.5 Library preparation for illumina sequencing

The ChIP-Seq libraries from two 8E stage Pol II ChIP and one bean stage Pol II ChIP were generated by using Apollo 324 System and PrepX ILM DNA Library Kit from IntergenX. After adaptor ligation, the input and ChIP DNA were enriched by PCR amplification using Phusion DNA Polymerase (New England Biolab) and Illumina universal PCR primer with the following PCR conditions: 30

seconds at 98°C, [10 seconds at 98°C, 30 seconds at 65°C, and 30 seconds at 72°C] for 5 cycles following 5 minutes at 72°C (14 uL adaptor ligated DNA, 15 uL NEBNext High-Fidelity 2X PCR master Mix, 1uL Universal PCR primer) (all reagents were included in NEBNext ChIP-Seq Library Prep Mix Set for Illumina [New England Biolab Cat# E6240S]). The enriched DNA was then purified using 30 uL (1:1 ratio of DNA volume and beads) of AMPure XP beads (Beckman Coulter) and selected 150–600 bp DNA by gel extraction. The size-selected DNA was amplified again using the same PCR conditions to amplify 5 more cycles (10 cycles total) for input libraries and 10 more cycles (15 cycles total) for ChIP libraries. After PCR, the libraries were purified by 50 uL (1:1 ratio of DNA volume and beads) and AMPure XP beads and eluted with 20 uL of TE buffer. 1 uL of each library was applied to measure the concentration using a Qubit dsDNA assay kit (Invitrogen). 1 ng of DNA from each library was checked by a Bioanalyzer (Agilent Technologies). Input and ChIP libraries were pooled such that they each had the same amount of molecules and expected for obtaining the similar number of reads. The Illumina sequencing was performed with 50 nt paired-end reads for two 8E stage samples and 100 nt paired-end reads for the bean stage sample.

3.3.6 ChIP-Seq data analysis

Raw data of ChIP-Seq from Illumina HiSeq 2000 were quality-checked using FastQC. The sequence reads were aligned to the ce10 version of the *C.*

elegans genome, obtained from the UCSC Genome Bioinformatics Site. Further filtering was performed to obtain uniquely mapped reads and remove duplicates.

To determine 5' poised Pol II, the poising index (PI) was calculated by the ration of mean ChIP-Seq reads in the promoter divided by the mean ChIP-Seq reads in the gene body for genes >1kb. Promoter Pol II signal was determined by Pol II peak found at 50 bp downstream or 250 bp upstream of a newly annotated TSS (Chen et al. 2013; Saito et al. 2013). A P-value of 0.05 was applied as a cutoff to define Pol II peaks.

3.3.7 RNA isolation and reverse transcription

For total RNA extraction, frozen embryo pellets were crushed by adding equal volume of glass beads (Sigma Cat # G8772-100G) and resuspended in Trizol Reagent (GibcoBRL), followed by chloroform extraction. RNA was precipitated with isopropanol and washed with 70% ethanol. Extracted RNA was treated with DNase before measuring concentration and checking the A260/A280 ratio using a Nanodrop spectrophotometer.

Reverse transcription was performed using a ProtoScript M-MuLV First Strand cDNA Synthesis Kit (New England Biolabs). 1 ug of extracted RNA and random hexamers were used to generate total cDNA. 4 ug of extracted RNA and strand specific primers were used to reverse transcribe strand specific transcripts. After reverse transcription, samples were treated with RNaseH and purified using a QIAquick PCR Purification Kit (QIAGEN).

3.4 Results

The motivation behind this study was to probe temporal Pol II occupancy during embryogenesis in order to understand if RNA Pol II is poised at early embryonic stages in *C. elegans* embryos and how the organ selector gene, PHA-4/FoxA, might play a role in regulating gene expression by affecting Pol II occupancy.

3.4.1 Different patterns of Pol II occupancy in *C. elegans* embryos

To study Pol II association during development, I analyzed genome-wide Pol II occupancy at different embryonic stages in wild type *C. elegans* embryos. We are specifically interested in the dynamics of Pol II occupancy at pharyngeal targets that are temporally expressed during pharynx development. Therefore, I performed Pol II ChIP-Seq using synchronized wild type 8E (before the pharyngeal primordium is formed) and bean (after the pharyngeal primordium is formed) embryos. Taking advantages of the published modENCODE data, which contain genome-wide Pol II occupancy in mixed mid- to late embryos with most embryos older than the bean stage, I confirmed that my ChIP was working properly and also tracked Pol II occupancy at 3 distinct phases—early, mid-, and late of embryogenesis. The 8E and bean Pol II ChIP-Seq were performed with 50 nucleotide (nt) paired-end and 100 nt paired-end reads of Illumina sequencing, respectively.

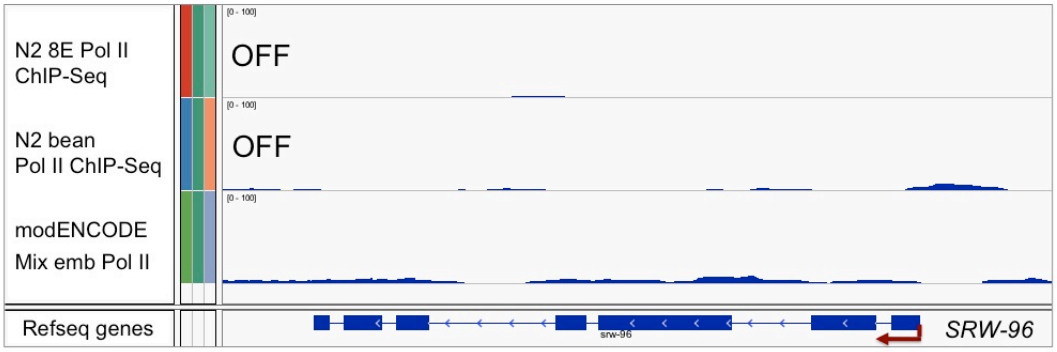
From both 8E and bean Pol II ChIP-Seq, I observed at least 4 different patterns of Pol II occupancy across the whole genome. (1) Silent or inactive genes had no or very few reads of Pol II signal across genes. (2) Active genes had Pol II occupancy across their entire genes. (3) Poised genes had Pol II enriched at their 5' promoter proximal regions. (4) Internal or 3' poised genes had Pol II occupancy within gene bodies particularly at the 3' regions of genes (Figure 3.1). This latter category of gene differed from active genes in that there was little to no Pol II signal at the 5' end.

To localize Pol II within genes, we needed to map its position relative to the TSS. *C. elegans* is unusual because it employs trans-splicing to generate mature mRNAs (Blumenthal, 2012). Primary transcripts are often spliced to leader sequences that act as the first exon. This feature of worms makes it difficult to identify the 5' ends of genes, which are removed from immature transcripts. Recent studies from three labs have assigned real TSS regions to *C. elegans* genes by either employing GRO-Seq with an immunoprecipitation step for the capped 5' mRNA (Kruesi et al., 2013) or by isolating unspliced RNA from embryos (Chen et al., 2013) or by nuclear isolation of cooled animals in which the splicing machinery is sensitive to cold temperature (Saito et al., 2013). To identify genes with poised Pol II, we applied the newly assigned TSS data (Chen et al*.,* 2013; Saito et al., 2013; Kruesi et al., 2013) to calculate the Pausing Index, in collaboration with Anat Burger and Erel Levine in the Harvard Physics Department. The Pausing Index is defined as the ratio of the ChIP-Seq signal in the TSS divided by the mean ChIP-Seq signal in the body of the gene (see

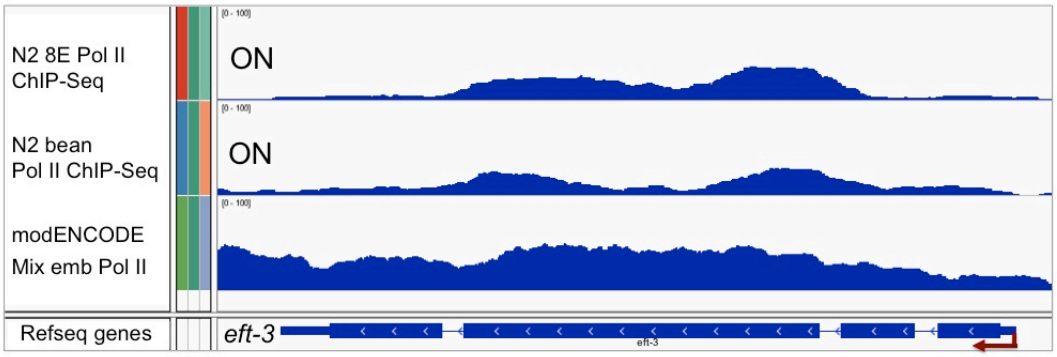
Figure 3.1

Different patterns of Pol II occupancy in *C. elegans* embryos. Pol II ChIP-Seq performed by Illumina sequencing indentified 4 different patterns of Pol II occupancy in *C. elegans* 8E and bean embryos. (A) No or very low Pol II signal was detected at silent genes, such as *srw-96*. *srw-96* is not expressed at 8E and bean stages (see Figure 3.2). (B) Pol II signal across the body of genes for active genes, such as *eft-3*. *eft-3* is actively expressed at both 8E and bean stages. (C) Pol II was enriched at 5' promoter proximal regions of poised genes, such as *ceh-22*. *ceh-22* is expressed from the bean stage onwards (Figure 3.2). (D) Pol II was enriched within gene bodies or at the 3' end of genes, such as *myo-2*. *myo-2* is not active until the 2-fold stage. modENCODE Pol II ChIP-Seq was from www.modencode.org (Zhong et al., 2010). Y axis indicated 3 Pol II ChIP-Seq performed in embryos at different stages, the 8E, bean and mixed-late embryos. ON: gene expression is active. OFF: gene expression is silent.

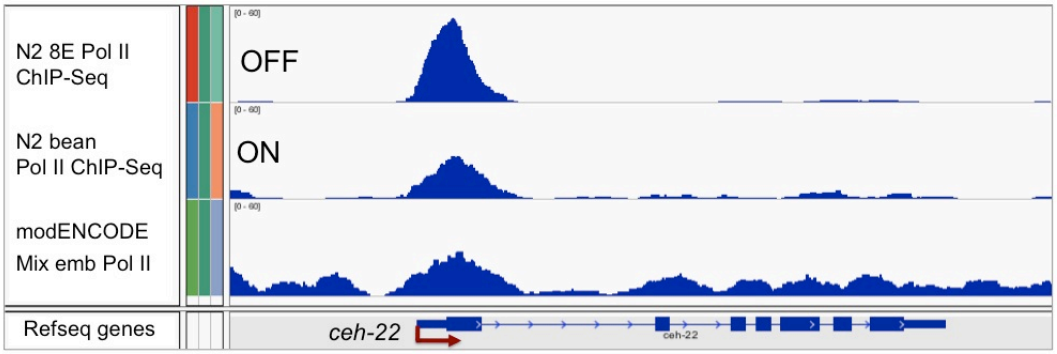
A **Silent Gene**



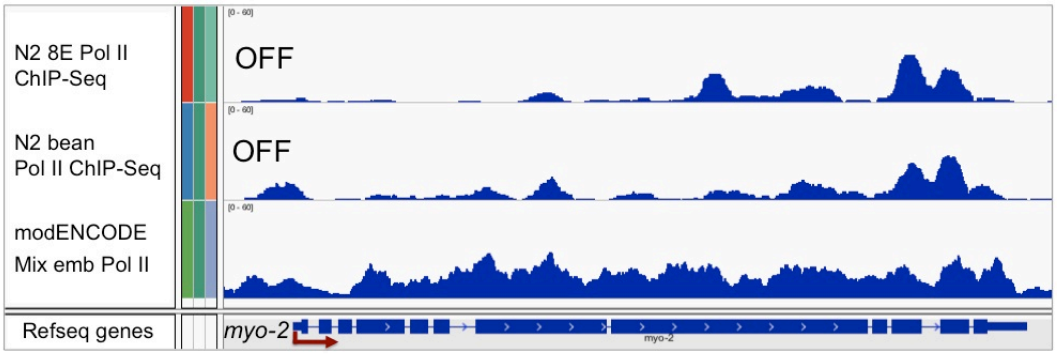
B **Active Gene**



C **Poised Gene**



D **Gene with internal Pol II peak**



materials and methods). Using a statistical cut-off of 0.05, we identified 4.6% (1069/23410) and 7.2% (1685/23410) of genes were poised at the 8E and bean stages, respectively. In comparison to our results, data from modENCODE and the Meyer and Lis labs identified 2% and 0.38% of poised genes in mixed-stage embryos. GRO-Seq results from the Meyer lab specifically identified paused genes in contrast to modENCODE Pol II ChIP-Seq that determined all forms of Pol II similar to our approach. The difference of statistical cut-off also affects the analysis of poising. We will verify whether we applied a reasonable cut-off by checking the Pol II patterns at genes excluded from a stringent cut-off.

I concentrated on analyzing poised Pol II on 161 pharyngeal genes previously identified (Gaudet et al., 2004). Using the same threshold, 10.5% of pharyngeal genes were poised at the 8E and bean embryos. Among the poised pharyngeal genes, 12% of genes were specifically poised at the 8E and 35% at the bean stage. This result implies that the regulation of poised Pol II varies with the developmental stages in *C. elegans* embryos. The degree of poised genes observed for the pharynx is a little higher than the genome as a whole, suggesting features of the pharynx may be optimal for poising. Nevertheless, it is clear that many genes do not show obvious poising, either because they are never poised or because I did not test the right stage of embryogenesis.

3.4.2 The dynamics of Pol II occupancy during development reflect the regulation and the status of gene expression

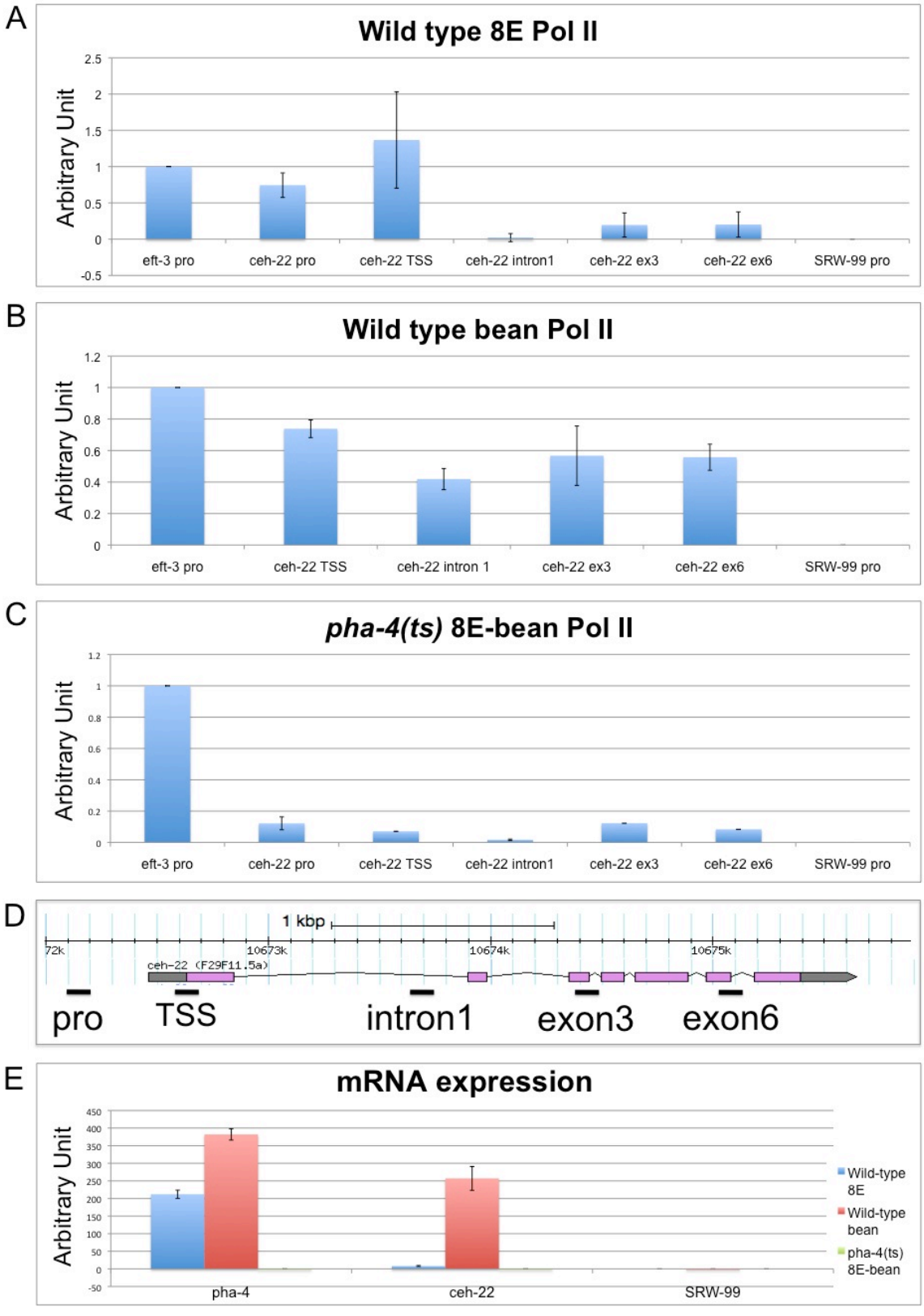
To confirm our genome-wide ChIP-Seq result of genes exhibiting poised Pol II, I chose five pharyngeal targets in the "poised" category and examined them by Pol II ChIP-qPCR from 8E embryos. I chose poised pharyngeal genes that exhibited different degrees of poised Pol II (we referred to it as poising rank) to test the threshold of our analysis. At the 8E stage, four of five poised targets, including high and low ranks of poising, shown in ChIP-Seq indeed exhibited enriched 5' Pol II signal in comparison to the gene bodies by ChIP-qPCR (Figure 3.2 and 3.3). This result supports our analysis using a p value of 0.05 as a cut-off. A widely expressed gene, *inx-3*, did not show 5' enriched Pol II in ChIP-qPCR. It is possible that inx-3 is not poised in all cell types as substantial Pol II signal was detected in the coding region of *inx-3* by ChIP-qPCR. The rest of four poised genes were specifically expressed in the pharynx determined by in situ hybridization (Tabara et al., 1996). Overall, the result of ChIP-qPCR is consistent with our finding from ChIP-Seq.

ChIP-qPCR also confirmed the pattern of internal and 3' poised Pol II at *myo-2* identified from ChIP-Seq. This result provides a validation of our Pol II ChIP-Seq analysis.

By comparing the patterns of Pol II occupancy to the gene expression, we found that the four different Pol II patterns reflected on the status of gene activity. Genes without or with very low reads of Pol II signals were transcriptionally silent or only expressed in a few cells. An example of this configuration is *srw-96*,

Figure 3.2

Dynamic Pol II occupancy at *ceh-22* loci reflects mRNA expression of *ceh-22*. (A) Pol II ChIP-qPCR of wild type 8E embryos with 5' enriched Pol II at the *ceh-22* promoter proximal region. (N > 3) (B) Pol II ChIP-qPCR in wild type bean stage embryos with Pol II occupancy within the *ceh-22* gene body. (N = 3) (C) Pol II ChIP-qPCR at *ceh-22* in *pha-4(ts)* mutant 8E-bean embryos. (N = 3) (D) qPCR primers target regions at *ceh-22*. (E) RT-qPCR detected *pha-4* and *ceh-22* mRNA in wild type embryos (blue, red panels) but not in *pha-4* mutants (green) (N = 2). (Pol II at *eft-3 / gene* promoter was a positive control for Pol II ChIP-qPCR and *srw-99* promoter was a negative control). Pol II occupancy at different *ceh-22* genomic regions was normalized to *eft-3* (set to 1) and *srw-99* gene (set to 0). *pha-4* mRNA expression served as a positive control for both RT-qPCR and an indication for embryonic stages since *pha-4* expression is increased from 8E to bean and eliminated in *pha-4(ts)* at the restrictive temperature.)


A
Wild type 8E Pol II
Arbitrary Unit
2.5
2
1.5
1
0.5
0
-0.5
eft-3 pro
ceh-22 pro
ceh-22 TSS
ceh-22 intron1
ceh-22 ex3
ceh-22 ex6
SRW-99 pro
B
Wild type bean Pol II
Arbitrary Unit
1.2
1
0.8
0.6
0.4
0.2
0
eft-3 pro
ceh-22 TSS
ceh-22 intron 1
ceh-22 ex3
ceh-22 ex6
SRW-99 pro
C
pha-4(ts) 8E-bean Pol II
Arbitrary Unit
1.2
1
0.8
0.6
0.4
0.2
0
eft-3 pro
ceh-22 pro
ceh-22 TSS
ceh-22 intron1
ceh-22 ex3
ceh-22 ex6
SRW-99 pro
D
1 kbp
72k
10673k
10674k
10675k
ceh-22 (F29F11.5a)
pro
TSS
intron1
exon3
exon6
E
mRNA expression
Arbitrary Unit
450
400
350
300
250
200
150
100
50
0
-50
pha-4
ceh-22
SRW-99
Wild-type 8E
Wild-type bean
pha-4(ts) 8E-bean

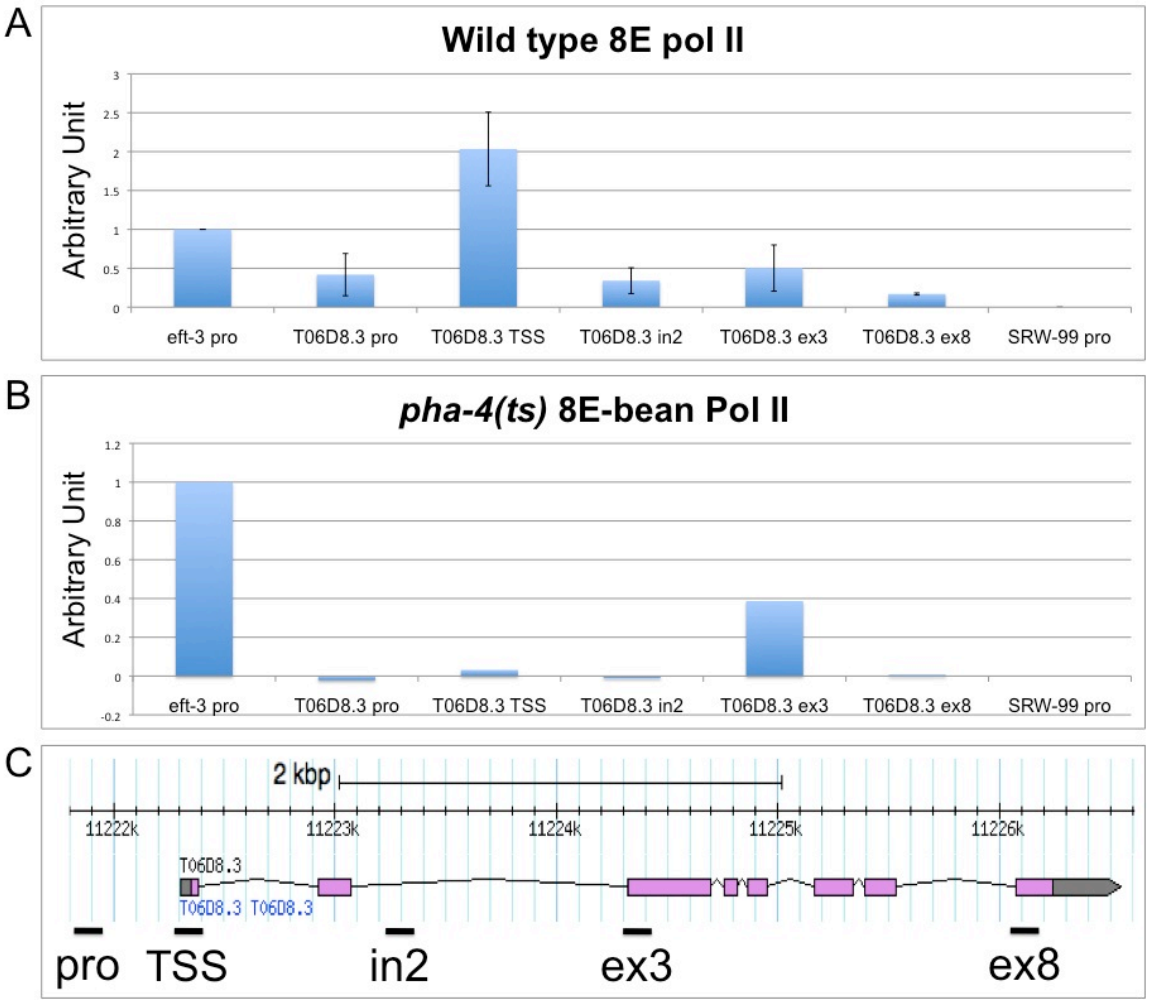


Figure 3.3

PHA-4 activity is required for Pol II occupancy at the TSS *T06D8.3*. (A) Pol II ChIP-qPCR in wild type 8E embryos with 5' enriched Pol II at the *T06D8.3* TSS (N = 2). (B) Pol II ChIP-qPCR in *pha-4(ts)* 8E-bean mutant embryos with decreased Pol II occupancy at the *T06D8.3* TSS (N = 1). (C) qPCR primers target regions at *T06D8.3*.

a serpentine receptor specifically activated in few neurons, which showed no detectable mRNA expression in embryos (Levin et al., 2012) (Figure 3.1A). Actively or ubiquitously expressed genes, such as housekeeping genes *eft-3* (translation elongation factor), *taf-1* (TATA-binding protein associated factor), and *his-72* (H3 histone), showed Pol II occupancy across the whole genes (Figure 3.1B). None of these genes are expressed selectively in the pharynx.

Genes that exhibited 5' enriched Pol II at the 8E and/or the bean stages were poised for activation later, including *ceh-22* and *T06D8.3* (Figure 3.1C, Figure 3.2, and Figure 3.3). From RT-qPCR analysis, *ceh-22* mRNA was not detected at the 8E stage but was increased significantly at the bean stage, in agreement with published studies (Figure 3.2) (Kuchenthal et al., 2001). I observed poised Pol II in the 8E embryos prior to activation and to a lesser degree at the bean stage, when *ceh-22* is transcribed. Pol II occupancy at *ceh-22* TSS was decreased concomitant with more Pol II detected in the gene body at the bean stage, which dropped the poising rank of *ceh-22* from 90 (8E) to 307 (bean). This observation suggests that the poised Pol II at *ceh-22* was released from 5' region to gene body as transcription proceeding. The similar trend of Pol II dynamics was also shown at *tbx-2* (an early pharyngeal muscle gene that is expressed at the 12E stage) whose poising rank was descended from 54 (8E) to 756 (bean). Comparing the change in Pol II occupancy at the 8E and bean stages, these data suggest that Pol II is potentially poised at the 5' end of a subset of pharyngeal genes during early embryonic stages.

There are pharyngeal genes in the last category containing Pol II peaks within gene bodies and 3' regions of genes. One of them is *myo-2*, which is not produced until late embryogenesis between the 2-fold and the 3-fold stages (Tabara et al., 1996). I detected Pol II occupancy within the gene body of the *myo-2* locus at the 8E stage, when no *myo-2* expression was detectable by RT-qPCR or using a transcriptional reporter (Figure 3.4; Gaudet and Mango, 2002). To examine if Pol II signal within *myo-2* gene body was associated with a transcriptional event, multiple primers were designed to hybridize with sense or antisense strands of *myo-2* RNA across the *myo-2* gene body. I used these primers to perform reverse transcription and analyzed the resulting cDNA by qPCR. This analysis revealed that antisense *myo-2* RNA was expressed at early embryonic stages (Figure 3.4), which suggests that the Pol II signal within the *myo-2* gene body was responsible for the *myo-2* antisense transcription. A similar pattern of antisense transcription was also found at the *ZK816.4* locus, another locus that is also expressed at very late embryonic stages (the 3-fold stage) (Gaudet and Mango, 2002).

An appealing, if speculative, hypothesis to explain these data is that antisense transcription could block sense transcription at early embryonic stages and perhaps also maintain an opened chromatin environment to promote the level of expression at later stages. Alternatively, antisense transcription might reflect read-through transcription from the upstream gene, given the high density of genes within the *C. elegans* genome. To distinguish between these possibilities, the Pol II occupancy within the intergenic region between the 3' end

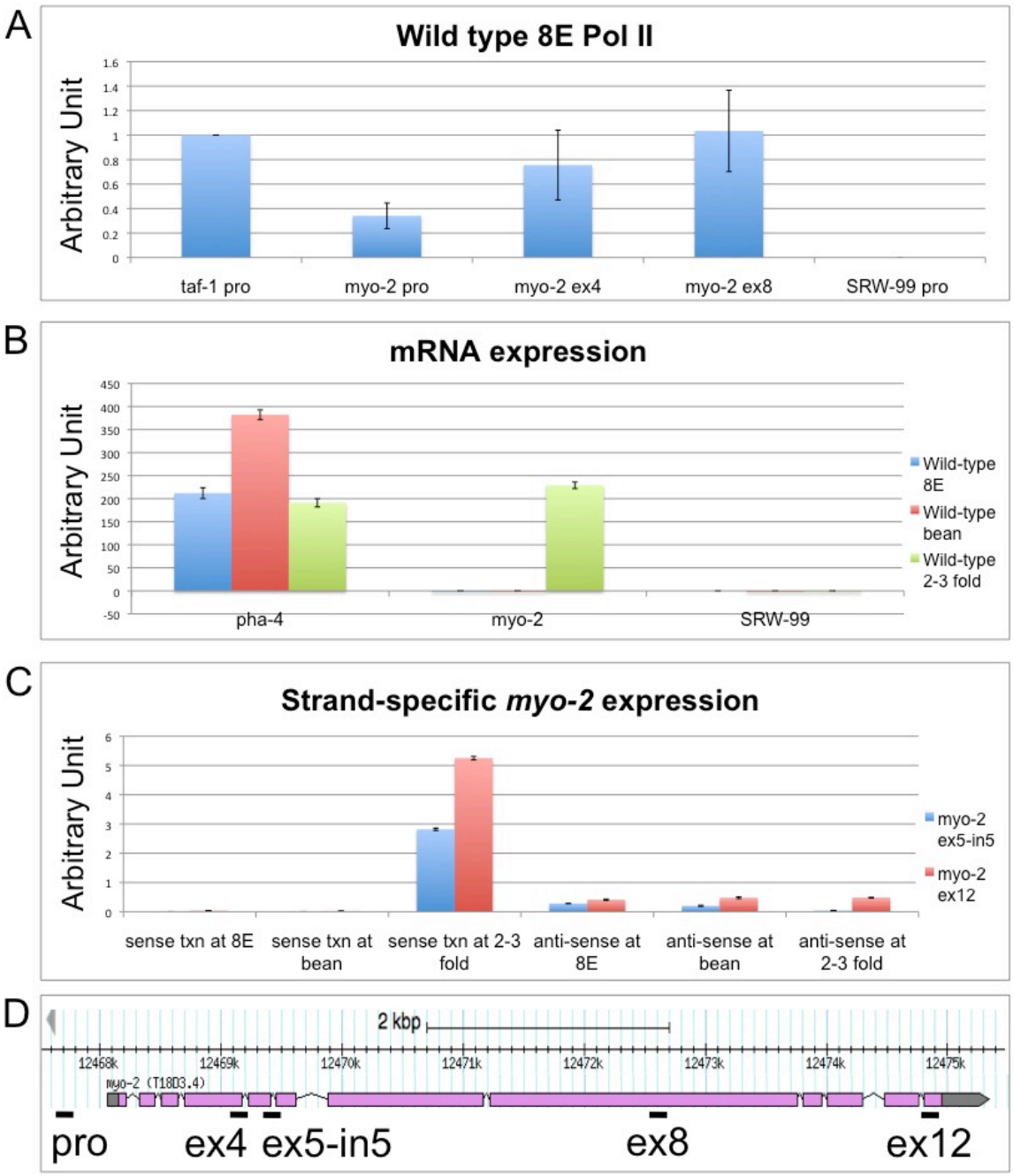


Figure 3.4

Pol II signal within gene body of *myo-2* before active transcription possibly accounts for the antisense transcription of *myo-2*. (A) Pol II ChIP-qPCR in wild type 8E embryos with Pol II within the *myo-2* gene body. (N > 3) (B) RT-qPCR to detect *myo-2* mRNA expression from 8E, bean to 2–3-fold. *myo-2* mRNA is not expressed until 2–3-fold. (N = 2) (C) Strand-specific RT-qPCR to detect strand specific *myo-2* expression. (N = 2) At 8E and bean, antisense but not sense transcription of *myo-2* was detected. (D) qPCR primers target regions at *myo-2*.

of *myo-2* and the 3' end of the downstream gene will be investigated by ChIP-qPCR.

In summary, I detected four configurations of Pol II on *C. elegans* genes that reflected their status of gene expression. Pol II occupancy was dynamically regulated during development. Combining Pol II ChIP-qPCR and RNA analysis, we discovered that pharyngeal gene expression was regulated in different ways, including Pol II poising and antisense transcription.

3.4.3 PHA-4 activity is required for Pol II occupancy at pharyngeal targets

PHA-4 activity is absolutely required to control pharyngeal gene expression during pharynx development. However, little is known about how pharyngeal transcription is regulated after PHA-4 binds to its targets. To test if *pha-4* activity affects Pol II occupancy at pharyngeal targets, I performed Pol II ChIP-qPCR in a *pha-4* temperature-sensitive mutant strain (see materials and methods). When *pha-4(ts)* mutant grows at the restrictive temperature, *pha-4* activity is eliminated and a pharynx fails to form (Kiefer et al., 2007). In *pha-4* mutant embryos at the 8E or bean stages, the 5' Pol II occupancy at *ceh-22* and *T06D8.3* (Figure 3.2C and Figure 3.3B) and *K10D3.4* (data not shown) was decreased when compared to wild type undergoing a similar temperature shift. This result indicates that PHA-4 activity is required for Pol II occupancy at pharyngeal targets. PHA-4 likely affects the recruitment of Pol II at pharyngeal promoters.

3.5 Discussion

3.5.1 Comparison of our finding to recently published Pol II studies

By performing Pol II ChIP-Seq in synchronized early (8E) and mid (bean) stage embryos, I identified four different patterns of Pol II occupancy. These patterns include a 5' enriched Pol II occupancy similar to the pattern of poised Pol II found in *Drosophila* embryos and mammalian cells (Nechaev et al., 2010; Rahl et al., 2010; Zeitlinger et al., 2007). Recently, several studies of genome wide Pol II occupancy in *C. elegans* have been published, including those by modENCODE (Baugh et al., 2009; Zhong et al., 2010; Kruesi et al., 2013). Poised Pol II in *C. elegans* was first reported in L1 larvae arrested by starvation (Baugh et al., 2009). Pol II was found to poise at approximately 10% of genes, which encoded regulators of growth and development. Further study from Kruesi and colleagues isolated short-capped RNA combining global run on sequencing (GRO-cap) to map transcriptional engaged Pol II in mixed-stage embryos and L1 larvae, which further confirmed that starvation can induce Pol II pausing (Kruesi et al., 2013). Many of the genes poised during starvation were activated in response to refeeding and subsequently exhibited increased Pol II occupancy over the coding regions. When comparing mRNA expression between poised and nonpoised genes, the expression of poised genes increased significantly during feeding recovery. This study indicates that poised Pol II reflects the memory of expression, but perhaps also the priming for future transcription in response to feeding.

Pol II ChIP-Seq performed in a *pha-4::gfp* integrated strain from modENCODE showed that fewer than 2% of genes exhibited 5' stalled Pol II in mix-staged embryos and L1 larvae under normal feeding conditions. Of these stalled genes, around 85% in embryos and half in L1 larvae were bound by PHA-4, suggesting that PHA-4 may play a role in Pol II stalling (Zhong et al., 2010). In addition, over 50% of PHA-4 bound and stalled genes were stalled either in embryos or in L1 larvae, which indicates Pol II stalling is specific to stages. PHA-4 has been shown to bind to 20% of genes in the genome (Zhong et al., 2010) and differentially regulates the temporal expression of its targets (Gaudet and Mango, 2002; Gaudet et al., 2004). The stalled Pol II at a subset of PHA-4 targets might reflect the mechanism for controlling the expression of these genes. These data suggest that PHA-4 might modulate differential gene expression in response to developmental and environmental cues partly through regulating Pol II.

The level of poised Pol II is defined by the ratio of Pol II signal around the 5' TSS to the average signal in the gene body. Therefore, precise assignment of the TSS of each gene is important to determine the poised genes. As mentioned previously, the precise information of TSS for the *C. elegans* genome was not available until three recently published studies (Chen et al., 2013; Saito et al., 2013; Kruesi et al., 2013). The recently published GRO-cap, which recovered nascent transcripts prior to trans-splicing, mapped and assigned TSSs for the *C. elegans* genome (Kruesi et al., 2013). In addition, GRO-Seq also provided the direction of each transcript. This information is important for distinguishing read-

through versus the 3' enriched transcription signals from the adjacent genes when examining Pol II poising at gene dense regions. The precise TSS assignment of the *C. elegans* genome was not applied in modENCODE analysis. Moreover, GRO-Seq is specific to determine paused genes (which contain transcriptional engaged Pol II) in contrast to the Pol II ChIP-Seq performed from modENCODE and our study, which detected all forms of Pol II. These reasons might explain why only 0.38% of genes were considered to be paused in mixed-stage embryos reported from the recent GRO-Seq study (Kruesi et al., 2013) when compared to the 2% stalled genes identified from modENCODE (Zhong et al., 2010).

In our research, we employed the TSS information determined by GRO-cap (Chen et al., 2013; Saito et al., 2013; Kruesi et al., 2013) and found genes with poised Pol II that had not been identified previously. One reason for these additional genes could be our use of a synchronized population of embryos. It has been shown in arrested *C. elegans* L1 larvae that paused Pol II is quickly resolved in response to feeding within an hour and exhibited increased Pol II occupancy along gene bodies (Baugh et al., 2009). This observation indicates that the regulation of Pol II is dynamic. In mixed-stage embryos, the Pol II signal at specific loci is contributed from both early (undifferentiated cells; genes have not yet activated) and late (differentiated cells; genes are actively transcribed) embryos. With the fast development of *C. elegans* embryos, late embryos will contribute more Pol II signal due to the increase of cell numbers. Therefore, it is possible that the 5' poised Pol II signal in early embryos is masked by elongating

Pol II from older embryos. With this concern, our approach provides a temporal specificity in capturing Pol II dynamics corresponding to transcriptional activity during development.

3.5.2 Following the dynamic Pol II occupancy during development reveals different mechanisms of gene regulation

Tracking Pol II occupancy from early (8E), mid- (bean), and late (modENCODE mixed late embryos) embryogenesis reveals that Pol II occupancy changes dynamically. For example, *ceh-22* is highly paused at the 8E stage without detectable Pol II within its gene body. At the bean stage, the Pol II signal within the gene body had increased, indicating *ceh-22* was active for transcription (Figure 3.1C). It still exhibited higher levels of Pol II at the 5' end, suggesting that poising was not completely lost at later stages. A subset of mid-stage activated pharyngeal targets exhibited similar Pol II dynamics with promoter proximal poised Pol II at early embryonic stages and less at later embryogenesis (Table A.2). More interestingly, poised PHA-4 targets indentified from our research showed stage specificity in which 12% and 35% of poised pharyngeal genes were only poised at 8E and bean stages, respectively (Gaudet and Mango, 2002; Gaudet et al., 2004). Pharyngeal genes specifically poised at the 8E stage show significant increase of mRNA expression after the bean stage (Levin et al., 2012). In addition, two pharyngeal genes specifically poised at the bean stage, *T04B8.2* and *cdc-25.2*, are highly expressed during early embryogenesis, and their

expression is reduced after the 8E stage (Table 3.2) (Levin et al., 2012). This observation agrees with the previous reports that poised Pol II is associated with both the preparation of gene activation and the memory of gene expression (Baugh et al., 2009; Gaertner et al., 2012). These data support that Pol II poising might be a mechanism to control temporal pharyngeal gene expression.

Why might *C. elegans* embryos require poised Pol II? Embryogenesis in *C. elegans* is rapid with cell divisions occurring at periods less than an hour at early embryonic stages. Having Pol II poised at the 5' end of genes during embryogenesis might provide a means to switch gene expression on and off to execute developmental decisions.

3.5.3 PHA-4 activity is required for Pol II recruitment at pharyngeal promoters

Eukaryotic gene expression can be regulated at the level of Pol II recruitment or alternatively by the release of paused Pol II for productive elongation. To understand how PHA-4 regulates pharyngeal gene expression, I analyzed the Pol II occupancy at poised pharyngeal promoters in *pha-4* mutant embryos. If PHA-4 regulates pharyngeal expression by Pol II recruitment, I expected a decrease in Pol II occupancy at promoter proximal regions when PHA-4 activity was eliminated. Examining Pol II at poised pharyngeal promoters *ceh-22* and *T06D8.3*, in *pha-4* mutant embryos, I found a significant decrease in Pol II occupancy at the TSS (Figure 3.2C and 3.3B). These data imply that PHA-4 activity affects the recruitment of Pol II at promoter proximal regions. I note that

this does not rule out an additional role for PHA-4 in the release of Pol II. There is precedence for this idea since in yeast the related fkh1 factor is thought to regulate transcriptional elongation by recruiting p-TEFb (Morillon et al., 2003). Other selector genes function at multiple stages of the transcription cycle. For example, MyoD controls over 300 myogenic target genes both by facilitating PIC assembly at promoters and by recruiting p-TEFb to phosphorylate Pol II and promote transcriptional elongation (Deato et al., 2008; Giacinti et al., 2006). PHA-4 functions as a pioneer factor that binds to pharyngeal targets long before gene activation to induce chromatin decompaction (Fakhouri et al., 2010; Friedman and Kaestner, 2006; Zaret et al., 2008). The open chromatin created after PHA-4 binding to its targets may facilitate Pol II assembly at promoter proximal regions or vice versa.

Growing *pha-4(ts)* at the restrictive temperature resulted in eliminating *pha-4* expression and pharyngeal gene expression (Figure 2.3E). Loss of PHA-4 activity early in development makes it difficult to examine PHA-4 function at both the recruitment and the release of Pol II. A follow-up experiment will be to examine Pol II dynamics at pharyngeal genes in the embryos undergoing a temporal increase and decrease in PHA-4 activity by heat shock and RNAi at different stages. As shown in Chapter 2, reduced PHA-4 level by RNAi resulted in delayed *Pceh-22*::GFP expression. What is the pattern of Pol II occupancy at *ceh-22* when PHA-4 level is low but can still activate *ceh-22* expression? These experiments will provide insights of PHA-4 regulated Pol II occupancy during transcriptional cycles.

In summary, my study provides a temporal resolution of Pol II occupancy during *C. elegans* embryogenesis. Our finding also gives a taste of how organ-specific poised genes were set up by a selector gene. How PHA-4 primes gene expression by regulating Pol II in a temporal manner and how genes are poised in general in *C. elegans* in comparison to other organisms will be the future focus.

3.6 References

Adkins, N.L., Hagerman, T.A., and Georgel, P. (2006). GAGA protein: a multi-faceted transcription factor. Biochem. Cell Biol. *84*, 559–567.

Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Mol. Cell *8*, 327–337.

Baugh, L., Demodena, J., and Sternberg, P. (2009). RNA Pol II accumulates at promoters of growth genes during developmental arrest. Science *324*, 92–94.

Blumenthal, T. (2012). Trans-splicing and operons in C. elegans. WormBook, 1–11.

Carroll, J., Liu, X., Brodsky, A., Li, W., Meyer, C., Szary, A., Eeckhoute, J., Shao, W., Hestermann, E., Geistlinger, T., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33–43.

Chen, R., Down, T., Stempor, P., Chen, Q., Egelhofer, T., Hillier, L., Jeffers, T., and Ahringer, J. (2013). The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. Genome Res. *23*, 1339–1347.

Cirillo, L., Lin, F., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Deato, M., Marr, M., Sottero, T., Inouye, C., Hu, P., and Tjian, R. (2008). MyoD targets TAF3/TRF3 to activate myogenin transcription. Mol. Cell *32*, 96–105.

Eberhardy, S.R., and Farnham, P.J. (2002). Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. J. Biol. Chem. *277*, 40156–40162.

Fakhouri, T., Stevenson, J., Chisholm, A., and Mango, S. (2010). Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1001060

Forcales, S.V. (2012). The BAF60c-MyoD complex poises chromatin for rapid transcription. Bioarchitecture *2*, 104–109.

Friedman, J.R., and Kaestner, K.H. (2006). The Foxa family of transcription factors in development and metabolism. Cell. Mol. Life Sci. *63*, 2317–2328.

Fuda, N., Ardehali, M., and Lis, J. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep. *2*, 1670–1683.

Gaudet, J., and Mango, S. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821–825.

Gaudet, J., Muttumu, S., Horner, M., and Mango, S. (2004). Whole-genome analysis of temporal gene expression during foregut development. PLoS Biol. *2*.
doi: 10.1371/journal.pbio.0020352

Giacinti, C., Bagella, L., Puri, P., Giordano, A., and Simone, C. (2006). MyoD recruits the cdk9/cyclin T2 complex on myogenic-genes regulatory regions. J. Cell. Physiol. *206*, 807–813.

Gilmour, D. (2009). Promoter proximal pausing on genes in metazoans. Chromosoma *118*, 1–10.

Kaltenbach, L.S., Updike, D.L., and Mango, S.E. (2005). Contribution of the amino and carboxyl termini for PHA-4/FoxA function in Caenorhabditis elegans. Dev. Dyn. *234*, 346–354.

Kiefer, J., Smith, P., and Mango, S. (2007). PHA-4/FoxA cooperates with TAM-1/TRIM to regulate cell fate restriction in the C. elegans foregut. Dev. Biol. *303*, 611–624.

Kruesi, W., Core, L., Waters, C., Lis, J., and Meyer, B. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. eLife *2*.
doi: http://dx.doi.org/10.7554/eLife.00808

Kuchenthal, C.A., Chen, W., and Okkema, P.G. (2001). Multiple enhancers contribute to expression of the NK-2 homeobox gene ceh-22 in C. elegans pharyngeal muscle. Genesis *31*, 156–166.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol. Cell. Biol. *28*, 3290–3300.

Lee, D.K., Duan, H.O., and Chang, C. (2001). Androgen receptor interacts with the positive elongation factor P-TEFb and enhances the efficiency of transcriptional elongation. J. Biol. Chem. *276*, 9978–9984.

Leibovitch, B.A., Lu, Q., Benjamin, L.R., Liu, Y., Gilmour, D.S., and Elgin, S.C. (2002). GAGA factor and the TFIID complex collaborate in generating an open chromatin structure at the Drosophila melanogaster hsp26 promoter. Mol. Cell. Biol. *22*, 6148–6157.

Levin, M., Hashimshony, T., Wagner, F., and Yanai, I. (2012). Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev. Cell *22*, 1101–1108.

Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400–403.

Lupien, M., Eeckhoute, J., Meyer, C., Wang, Q., Zhang, Y., Li, W., Carroll, J., Liu, X., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958–970.

Mango, S. (2009). The molecular basis of organ formation: insights from the C. elegans foregut. Annu. Rev. Cell Dev. Biol. *25*, 597–628.

Mann, R.S., and Carroll, S.B. (2002). Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genet. Dev. *12*, 592–600.

Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N., and Mellor, J. (2003). Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. Science *300*, 492–495.

Morse, R. (2007). Transcription factor access to promoter elements. J. Cell. Biochem. *102*, 560–570.

Muse, G., Gilchrist, D., Nechaev, S., Shah, R., Parker, J., Grissom, S., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Nechaev, S., Fargo, D., Dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335–338.

Nojima, M., Huang, Y., Tyagi, M., Kao, H.-Y., and Fujinaga, K. (2008). The positive transcription elongation factor b is an essential cofactor for the activation of transcription by myocyte enhancer factor 2. J. Mol. Biol. *382*, 275–287.

Puri, P.L., Avantaggiati, M.L., Balsano, C., Sang, N., Graessmann, A., Giordano, A., and Levrero, M. (1997). p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription. EMBO J. *16*, 369–383.

Rahl, P., Lin, C., Seila, A., Flynn, R., McCuine, S., Burge, C., Sharp, P., and Young, R. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Saito, T., Hashimoto, S.-I., Gu, S., Morton, J., Stadler, M., Blumenthal, T., Fire, A., and Morishita, S. (2013). The transcription start site landscape of C. elegans. Genome Res. *23*, 1348–1361.

Satija, R., and Bradley, R. (2012). The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. Genome Res. *22*, 656–665.

Saunders, A., Core, L., Sutcliffe, C., Lis, J., and Ashe, H. (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. Genes Dev. *27*, 1146–1158.

Simone, C., Stiegler, P., Forcales, S.V., Bagella, L., De Luca, A., Sartorelli, V., Giordano, A., and Puri, P.L. (2004). Deacetylase recruitment by the C/H3 domain of the acetyltransferase p300. Oncogene *23*, 2177–2187.

Tabara, H., Motohashi, T., and Kohara, Y. (1996). A multi-well version of in situ hybridization on whole mount embryos of Caenorhabditis elegans. Nucleic Acids Res. *24*, 2119–2124.

Updike, D., and Mango, S. (2006). Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. PLoS Genet. *2*.
doi: 10.1371/journal.pgen.0020161

Wilkins, R.C., and Lis, J.T. (1997). Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. Nucleic Acids Res. *25*, 3963–3968.

Zaret, K., and Carroll, J. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zaret, K.S., Watts, J., Xu, J., Wandzioch, E., Smale, S.T., and Sekiya, T. (2008). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. Cold Spring Harb. Symp. Quant. Biol. *73*, 119–126.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nature Genet. *39*, 1512–1516.

Zhang, S., Li, X., Burghardt, R., Smith, R., 3rd, and Safe, S.H. (2005). Role of estrogen receptor (ER) alpha in insulin-like growth factor (IGF)-I-induced responses in MCF-7 breast cancer cells. J. Mol. Endo. *35*, 433–447.

Zhong, M., Niu, W., Lu, Z., Sarov, M., Murray, J., Janette, J., Raha, D., Sheaffer, K., Lam, H., Preston, E.*, et al.* (2010). Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1000848

Zhou, Q., Li, T., and Price, D. (2012). RNA polymerase II elongation control. Annu. Rev. Biochem. *81*, 119–143.

CHAPTER 4

SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS

4.1 Summary

The study described in this dissertation explored the molecular mechanisms of differential gene expression regulated by selector genes. Selector genes play an important role in specifying biological identity of cells, tissues, and organs during development, and their functions are often conserved among different organisms. The overview of Chapter 1 highlighted the ability of selector genes to orchestrate comprehensive transcriptional programs to fulfill the requirements for building specific functional structures, such as the pharynx in *C. elegans*. Molecular mechanisms of differential gene expression were reviewed and compared among different selector genes. Hierarchal activation of gene expression cascade is utilized to regulate the development of tissues and organs with less complexity, such as the intestine in *C. elegans,* which is composed of single cell type. Combinatorial regulation by selector genes and other transcription factors are incorporated into feed-forward circuits that lead to temporal and spatial gene expression patterns for other organs and tissues. This mechanism is a common strategy employed by many selector genes, including MyoD and PHA-4/FoxA. Here, I specifically focused on PHA-4/FoxA-regulated temporal pharyngeal gene expression. According to previously proposed models, PHA-4 mediates the onset of gene expression through binding affinity to its targets coupled to combinatorial regulation (Gaudet and Mango, 2002; Gaudet et al., 2004). Inputs from both PHA-4 affinity sites and temporal elements contribute to precisely time pharyngeal gene expression.

This thesis also considered the events downstream of selector genes, specifically the role of Pol II poising in regulating differential gene expression during development. Pol II poising is pervasive in flies, mice, and human cells but rare in worms. The similarity and difference of Pol II regulation between *Drosophila* and *C. elegans* was discussed to better understand Pol II pausing in regulating differential gene expression.

In Chapter 2, I concentrated on the PHA-4 affinity model for temporal control of pharyngeal expression. The affinity model predicts that pharyngeal expression should be sensitive to the level of PHA-4, including overexpression of PHA-4. This idea has never been tested. I tested the affinity model by investigating pharyngeal expression patterns with narrow (2–3-fold) or drastic (>10 fold) changes of PHA-4 levels. Dramatically increased or decreased PHA-4 levels affected the pharyngeal expression patterns. This result suggests that pharyngeal expression is sensitive to PHA-4 concentration but can maintain robust expression within a range of PHA-4 level fluctuation. This is in contrast to mammals, which are sensitive to a two-fold reduction in FoxA2 activity (Ang and Rossant, 1994; Kittappa et al., 2007; Weinstein et al., 1994). I further explored binding site affinity by examining PHA-4 binding to different affinity targets in vivo, using NSA and ChIP. This analysis revealed that PHA-4 bound to both high and low affinity targets hours before active transcription. In addition, the level of PHA-4 occupancy was affected by the affinity of PHA-4 binding sites within the target promoters. However, the effect of affinity on the level of PHA-4 binding was variable among different pharyngeal promoters, as we observed a greater

change of PHA-4 binding when the binding site affinity was altered within the *M05B5.2* promoter but not at the *ceh-22*. The drastic decrease in PHA-4 binding at the *M05B5.2* promoter with a down mutation of PHA-4 binding site is correlated with delayed onset of the GFP reporter driven by the same promoter construct. These results indicate that binding site affinity plays a role in regulating the level of PHA-4 occupancy at target promoters to modulate the onset of gene expression.

Sequence-specific transcription factors modulate specific transcriptional programs through directly or indirectly regulating RNA Pol II at target genes. As studies of poised Pol II in flies and mammals reveal its important role in regulating differential gene expression, Pol II poising in worms is still largely unknown. In Chapter 3, I examined dynamic Pol II occupancy during embryogenesis by performing Pol II ChIP-Seq in synchronized early (8E) and mid (bean) embryos. I identified genes that exhibited different patterns of Pol II occupancy that reflected their transcriptional status. Moreover, 4% and 7% of genes, including a subset of pharyngeal genes, showed promoter proximal poised Pol II at the 8E and bean stages, respectively, which suggests that Pol II poising is stage specific in *C. elegans* embryos. To test the idea that PHA-4/FoxA functions as a pioneer factor might be involved in establishing poised Pol II, I examined Pol II occupancy at poised pharyngeal genes in *pha-4(ts)* mutant embryos. Promoter occupancy of Pol II at poised pharyngeal genes was greatly decreased when PHA-4 was eliminated. This result indicates that PHA-4 activity is required to load Pol II at a subset of pharyngeal genes. I hypothesize that

PHA-4 primes gene activation partly through affecting temporal Pol II occupancy at pharyngeal promoters. This is exciting because I am not aware of papers showing a link between a selector gene and poised Pol II. Perhaps the decompaction of chromatin observed for FoxA factors relates to loading of Pol II. However, at this point we do not know if chromatin opening facilitates Pol II loading or if loading of Pol II destabilizes a nucleosome. This question will be the subject of future experiments.

4.2 Conclusions

Binding affinity between transcription factors and their recognition consensus sequences is used in many developmental contexts to regulate differential gene expression (Gaudet and Mango, 2002; Hollenhorst et al., 2009; Lam et al., 2008; Stathopoulos and Levine, 2004). Our study of PHA-4 affinity model in temporal gene regulation provides insights to understand the basic molecular mechanisms in regulating and maintaining a precise transcriptional program. First, PHA-4 binds to both high and low affinity targets prior to active transcription at early embryonic stages. Second, the binding site affinity impacted the level of PHA-4 occupancy at its target promoters. Third, the effect of affinity on PHA-4 binding was most pronounced at early embryonic stages. Fourth, PHA-4 levels normally increase during embryogenesis, and alterations in PHA-4 levels affected the onset of pharyngeal expression, especially at early embryonic stages from the 2E to 8E stages. These observations were reminiscent of the open chromatin configuration associated with developmental plasticity in early

embryos (Meister et al., 2011; Meshorer and Misteli, 2006; Yuzyuk et al., 2009). I hypothesize that the decompacted and accessible chromatin at early embryonic stages (Yuzyuk et al., 2009) permits PHA-4 binding to targets with different affinity sites even when PHA-4 expression level is still low at early stages. However, low affinity sites, coupled to low levels of PHA-4 at early stages, lead to low occupancy. Therefore, the affinity effect on the level of PHA-4 binding is emphasized at earlier stages compared to later stages, which explains the advanced onset of pharyngeal activation when PHA-4 was dramatically increased at early embryonic stages but not later stages. These data explore how PHA-4/FoxA as a pioneer factor broadly regulates pharyngeal gene expression through binding affinity with its targets.

Pol II poising at developmental regulators provides a strategy to set up developmental programs in the early *Drosophila* embryos (Chopra et al., 2011; Zeitlinger et al., 2007), but much less is known about the role of poising in *C. elegans* where poising is less pervasive. The Pol II ChIP-Seq study described in Chapter 3 provided evidence of poised Pol II in early *C. elegans* embryos. My study is similar to the previously characterized Pol II poising in *Drosophila* embryos where poised Pol II is specific to embryonic stages and dynamically regulated over time (Gaertner et al., 2012). Similarly, I observe loci with poised Pol II at the 8E stage but not the bean stage and vice versa. Other loci show poised Pol II both early and late.

Focusing on a subset of pharyngeal genes, I demonstrated that PHA-4/FoxA was required for Pol II occupancy at poised pharyngeal promoters. This

result suggests that other Fox factors might function similarly to mediate gene expression through regulation of Pol II dynamics and raises the question whether other selector genes also regulate poising. Overall, my work delineated the role of PHA-4 function in establishing a precise pharyngeal transcription during pharynx development.

4.3 Future directions

The analysis of PHA-4 binding to individual targets by NSA and ChIP provides a framework for studying transcriptional regulation. To gain a comprehensive understanding of PHA-4 orchestrated foregut formation, it would be insightful to examine the genome-wide temporal PHA-4 binding patterns at endogenous loci. I attempted this analysis but encountered difficulties with the antibodies used in ChIP. A genome-wide study of temporal PHA-4 binding could solidify and expand our understanding about binding site affinity involved in global gene regulation during organogenesis.

How does the level of PHA-4 occupancy at different affinity sites translate to expression onset? A possible mechanism is through controlling the recruitment of downstream factors required for pharyngeal activation, including cofactors, other transcription factors, and RNA Pol II. Our NSA studies showed that PHA-4 binding to its targets resulted in decompaction of the arrays (Fakhouri et al., 2010) and loading of the histone variant H2A.Z (Updike and Mango, 2006). Further studies combining PHA-4 binding affinity illustrated in Chapter 2 demonstrated that the level of PHA-4 occupancy was correlated with the

decompacted array sizes. Taking advantage of the temporal and single cell resolution of the NSA, we could study the affinity effect on the timing of downstream recruitment of other factors. For example, one could generate arrays carrying the *myo-2* promoter with different PHA-4 affinity sites and examine the patterns of CEH-22 binding.

One intriguing result from the NSA is that the PHA-4 affinity effect was different for *M05B5.2* and *ceh-22*. I observed a dramatic effect on the level of PHA-4 binding when the PHA-4 affinity site was mutated in the *M05B5.2* promoter but the effect was minor for the *ceh-22* promoter. Inspired by the Pol II study in Chapter 3, one possible explanation is the different patterns of Pol II occupancy at the *M05B5.2* and *ceh-22* promoters. Pol II was poised at *ceh-22* but not at *M05B5.2* from the 8E stage. In *Drosophila*, poised Pol II can function to maintain an open chromatin configuration (Gilchrist et al., 2010; Gilchrist et al., 2008). Perhaps, poised Pol II at the *ceh-22* promoter promotes chromatin opening in combination with PHA-4. This idea could be tested by examining the level of Ser5p Pol II (the transcriptional engaged Pol II) on the arrays carrying either *M05B5.2* or *ceh-22* promoter. From another angle, we could also examine the nucleosome profiles at endogenous *M05B5.2* and *ceh-22* loci. The dynamic genome-wide nucleosome profile during *C. elegans* embryogenesis is an on-going project being carried out by Huei-Mei Chen, a talented postdoctoral fellow in our lab. Her study will explore the connection between chromatin dynamics, PHA-4, and transcriptional regulation.

An appealing finding from the Pol II study is that PHA-4 might play a role in establishing poised Pol II at pharyngeal genes. Perhaps this is how PHA-4 functions as a pioneer transcription factor. We have tested the requirement of PHA-4 activity on Pol II occupancy at several poised pharyngeal genes. Further Pol II ChIP-Seq in *pha-4(ts)* embryos will confirm this argument.

Losing 5' poised Pol II at a subset of pharyngeal genes that is caused by the lack of PHA-4 activity during early embryogenesis makes it hard to interpret whether PHA-4 is also involved in regulating Pol II after its recruitment at promoters. We have shown that the perturbation of PHA-4 expression to increase or decrease PHA-4 levels resulted in shifting the temporal pharyngeal activation, including the *ceh-22,* which exhibited poised Pol II. Using *ceh-22* as an example to study PHA-4 and the dynamic changes of Pol II occupancy would further explore the mechanisms of PHA-4 controlled temporal gene expression.

4.4 References

Ang, S.L., and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. Cell *78*, 561–574.

Chopra, V., Hendrix, D., Core, L., Tsui, C., Lis, J., and Levine, M. (2011). The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. Mol. Cell *42*, 837–844.

Fakhouri, T., Stevenson, J., Chisholm, A., and Mango, S. (2010). Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. PLoS Genet. *6*.
doi: 10.1371/journal.pgen.1001060

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep. *2*, 1670–1683.

Gaudet, J., and Mango, S. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821–825.

Gaudet, J., Muttumu, S., Horner, M., and Mango, S. (2004). Whole-genome analysis of temporal gene expression during foregut development. PLoS Biol. *2*. doi: 10.1371/journal.pbio.0020352

Gilchrist, D., Dos Santos, G., Fargo, D., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540–551.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev. *22*, 1921–1933.

Hollenhorst, P.C., Chandler, K.J., Poulsen, R.L., Johnson, W.E., Speck, N.A., and Graves, B.J. (2009). DNA specificity determinants associate with distinct transcription factor functions. PLoS Genet. *5*, e1000778.

Kittappa, R., Chang, W.W., Awatramani, R.B., and McKay, R.D. (2007). The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age. PLoS Biol. *5*, e325.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246–250.

Meister, P., Mango, S.E., and Gasser, S.M. (2011). Locking the genome: nuclear organization and cell fate. Curr. Opin. Genet. Dev. *21*, 167–174.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell. Biol. *7*, 540–546.

Stathopoulos, A., and Levine, M. (2004). Whole-genome analysis of Drosophila gastrulation. Curr. Opin. Genet. Dev. *14*, 477–484.

Updike, D., and Mango, S. (2006). Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. PLoS Genet *2*. doi: 10.1371/journal.pgen.0020161

Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., and Darnell, J.E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell *78*, 575–588.

Yuzyuk, T., Fakhouri, T., Kiefer, J., and Mango, S. (2009). The polycomb complex protein mes-2/E(z) promotes the transition from developmental plasticity to differentiation in C. elegans embryos. Dev. Cell *16*, 699–710.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512–1516.

APPENDIX

Table A.1
Sequences of primers used for ChIP-qPCR and RT-qPCR.

Primer name	Sequences	Note
taf-1 pro-F	TTTTTCAGCGAAATGATTGTAAA	
taf-1 pro-R	CAACCAACTCCGGAAAAACTAA	
eft-3 pro-F	AGCGTTTTTCCTGTTCTCACTGTTT	
eft-3 pro-R	GAGTGCGGACGGTAGAGAGAATAAA	
ceh-22 pro-F	GCTTTCGGAGCATTAACCTAGA	
ceh-22 pro-R	TGAGCATCTCGGCTAACAATAA	
ceh-22 TSS-F	AGCAACATGGATTGTCTACAAGTG	
ceh-22 TSS-R	TACTTGTTATGAAGTGCCGATTGT	
ceh-22 intron1-F	TGAAACGGGATGAAAATATCCT	
ceh-22 intron1-R	TTTTCTTTGAAAATGGGGAGAA	
ceh-22 ex3-F	CTTCAAACGGAATTGCTGGTAAGTTTTTTC	
ceh-22 ex3-R	GATCAGGTAGAGTGTACATGGAACCTACA A	
ceh-22 ex6-F	AAGTGCCAATTCAGGATACCTACC	
ceh-22 ex6-R	GTCATGTAGGAAGAAGCTGCTGAA	
SRW-99 pro-F	TCAAAATGTTCCCACGTCTATAAAACTTAC	
SRW-99 pro-R	AAAAAGAGCATTTTGCAATACGTTAGAG	
T06D8.3 pro-F	CTAGTTATACCGTCCGCCTACCTGT	
T06D8.3 pro-R	ATGCTTAACTGTTCAGGGGGAGA	
T06D8.3 TSS-F	GCAACGACATGTCCGTGCCAGCTTCG	
T06D8.3 TSS-R	ATGTAAAGAAATCTGTCGCTTCGAG	
T06D8.3 in2-F	GGTACTACTTGAGGTAAGCCTTTGACA	
T06D8.3 in2-R	CTCGCCGAGCAAATCATAGTACAAAC	
T06D8.3 ex3-F	TTTTGTCACTATCGCTGAGGCTTAC	
T06D8.3 ex3-R	CATCTCGACAGTAAAATACCCGATG	
T06D8.3 ex8-F	AATCCTTCCATAGAACCATCAGTCC	
T06D8.3 ex8-R	ATGCACCTAGTACCGATATCCATCA	
myo-2 pro-F	ATGGCAGGAAGAGCACTTTG	
myo-2 pro-R	ATGCAGAGAGGCAGACATCC	
myo-2 ex4-F	GGAGAATCTGGTGCAGGAAA	
myo-2 ex4-R	GCTCCGAAGGTCTCTTGTTG	
myo-2 ex8-F	CACCAACAAACCATCGACAG	Used for both ChIP-qPCR and RT-qPCR
myo-2 ex8-R	CGGCTTCCAACTTCTTCTTG	
pha-4_qPCR_Fwd	CGGCTGTTAATCACAGTCAACCTACTTCAG	
pha-4_qPCR_Rev	CCGAACTGTAGAGGTAAGGAGACGC	
ceh-22 ex1-ex2 F	CTTCATAACAAACTCGAAGCTAAATGGG	
ceh-22 ex2-ex3 R	GGAGTAGCAGAATATCCTGCAAGTAATGG	
SRW-99 ex1-ex2 F	TTGTCAATGACATTGCCAATTACGCGC	
SRW-99 ex2 R	GCAATTCCAATTAATATCACGTTTGTAGAA G	

Table A.1 continued.

P2	GCTTCAGATCAAGGTGACC	
myo-2 AS-P2-ex5-F	GCTTCAGATCAAGGTGACCCTTCTCCGATTACTTGCCAAAC	
myo-2 AS-in5-R	CTGCATTTCGTCATGTTTTGAT	
myo-2 AS-P2-ex12-F	GCTTCAGATCAAGGTGACCAGAAGCAACTCGAGGAAGCTGT	
myo-2 AS-ex12-R	GGGAAAGTTGGACGGTTCTGTA	
myo-2 S-ex5-F	GTTTGGCAAGTAATCGGAGAAG	
myo-2 S-in4-R	CGAATGTTTGACCTGTATATAT	
myo-2 S-ex12-F	ACAGCTTCCTCGAGTTGCTTCT	
myo-2 S-ex12-R	CAGAGTGACTTGGAAACCGAGA	

Table A.2
The list of poised pharyngeal genes.

Gene	8E poising rank	bean poising rank	Encoded protein	Expression (Levin *et al.* 2012)
Poised at the 8E				
alp-1	219		α-actin associated protein	From the bean stage
F30H5.3	919		unknown	From the bean stage
Poised at the 8E and bean				
Inx-3	84	276	innexin protein	Constantly express
ceh-22	90	307	NKX homeobox transcription factor	From the bean stage
ain-1	745	655	GW182 family protein, involve in miRNA-mediated gene silencing	From the pre-bean stage
D1054.9	201	814	unknown	From the comma stage
ces-2	455	907	bZIP transcription factor	From the pre-bean stage
T06D8.3	338	1082	unknown	From the bean stage

Table A.2 continued.

tbx-2	54	756	T-box transcription factor	From the 12E stage
fbxc-23	576	737	F-box c protein	From the pre-bean stage
inos-1	1059	1561	myo-inositol-1-phosphate synthase	Constantly express
Poised at the bean				
gmeb-2		21	GMEB (Glucocorticoid Modulatory Element Binding protein) transcriptional regulator homolog	From the comma stage
T04B8.2		263	F-box protein	Express early from the 2E, expression is dropped at the comma stage
cdc-25.2		593	Cdc25 phosphatase	Express early from the 2E and expression is decreased after the 8E
K10D3.4		615	unknown	From the bean stage
T28B8.1		1482	unknown	From the bean stage
T09B4.5		1634	unknown	From the bean stage